Ghent University – Faculty of Sciences – Dept. Biotechnology and Bioinformatics

VIB – Dept. Plant Systems biology

Natural Variation in the Mild Drought Response of *Arabidopsis thaliana* Leaves

Pieter Clauw

2014-2015

Promoter: Prof. dr. Dirk Inzé

Dissertation submitted in fulfillment
of the requirements to obtain the degree
of Doctor of Philosophy (Ph.D.) in Sciences:
Biochemistry and Biotechnology

- Maps only become interesting at their edges -

Examination Board

CHAIR

Prof. dr. Tom Beeckman
Ghent University – Department of Plant Biotechnology and Bioinformatics
VIB – Department of Plant Systems Biology

SECRETARY

Prof. dr. Dirk Inzé
Ghent University – Department of Plant Biotechnology and Bioinformatics
VIB – Department of Plant Systems Biology

MEMBERS

Dr. Frederik Coppens
Ghent University – Department of Plant Biotechnology and Bioinformatics
VIB – Department of Plant Systems Biology

Dr. Nathalie Gonzalez
Ghent University – Department of Plant Biotechnology and Bioinformatics
VIB – Department of Plant Systems Biology

Dr. Christian Hermans*
Université Libre de Bruxelles – Laboratory of Plant Physiology and Molecular Genetics

Dr. Arthur Korte*
Gregor Mendel Institute of Molecular Plant Biology – Population Genetics

Prof. dr. Steven Maere*
Ghent University – Department of Plant Biotechnology and Bioinformatics
VIB – Department of Plant Systems Biology

* Member of the reading committee

Contents

List of Abbreviations

ABA	Abscisic Acid
ABF	ABRE-Binding Factor
ABRE	ABA Responsive Elements
ACC	Aminocyclopropane-1-Carboxylic Acid
AHL	AT-Hook Motif Nuclear-Localized Protein
AIL	Advanced Intercross Line
AMPRIL	Arabidopsis Multiparent RIL
APC	Anaphase-Promoting Complex
AREB	ABRE-Binding Protein
bHLH	Basic Helix-Loop-Helix
CAST	Cluster Affinity Search Technique
CDK	Cyclin-dependent Kinase
Col	Columbia
CPK	Calcium Dependent Protein Kinase
CSLC	Cellulose Synthase Like
CYC	Cyclin
CYP	Cytochrome P450
DAS	Days After Stratification
DRE	Dehydration Response Element
DREB	DRE Binding Protein
eGWAS	expression GWAS
eQTL	expression QTL
ERD	Early Responsive to Dehydration
ERF	Ethylene Response Factor
EXL	Exordium-Like
EXP	Expansin
FAD	Falvine Andenine Dinucleotide
FDR	False Discovery Rate
FLA	Fasciclin-Like Arabinogalactan
GA	Gibberellic Acid
GA20OX	GA20 Oxidase
GA2OX	GA2-Oxidase
GO	Gene Ontology

GWAS	Genome Wide Association Study/Mapping
G×E	Genotype by Environment
HB	Homeo Box
INDEL	Insertion/Deletion
JA	Jasmonic Acid
JAW	Jagged And Wavy
LD	Linkage Disequilibrium
LEA	Late Embrtogenesis Abundant
lncRNA	Long Non-Coding RNA
LOX	Lipoxygenase
MAGIC	Multiparent Advanced Genetic Intercross
miRNA	Micro RNA
MLP	Major Latex Protein-Like Protein
MPK	Mitogen-Activated Protein Kinase?
MTMM	Multi-Trait Mixed Model
NAC	NAC Domain containing Protein
NCED	Nine-Cis-Epoxycarotenoid Dioxygenase
NLP	Nodule-Like Inception Protein
P	P-value
P5CS1	Delta1-Pyrroline-5-Carboxylate Synthase
PLC8	Phosphatidylinositol-Specific Phospholipase C8
PP2C	Protein Phosphatase 2C
PRA	Projected Rosette Area
PSK	Phytosulfokine
PYL	PYR-Like
PYR	Pyrabactin Resistant
QTL	Quantitative Trait Locus
r	Correlation coefficient
R-gene	Resistance Gene
RCAR	Regulatory Component of ABA Receptor
RIL	Recombinant Inbred Line
RLP	Receptor Like Protein
ROS	Reactive Oxygen Species
SA	Salicylic Acid
SDH	Succinate Dehydrogenase
SE	Standard Error

SIP	Small and Basic Intrinsic Protein
siRNA	Small Interfering RNA
SNP	Single Nucleotide Polymorphism
SnRK	Sucrose Non-Fermenting1-Related Kinases
STG	Stress Tolerance Gene
TCP	Teosinte Branched 1/Cycloidea/Proliferating Cell Factor 1
TE	Transposable Element
TRE	Trehalase
UFO	Unusual Floral Organs
WIWAM	Weighing Imaging and Watering Automated Machine
WT	Wild Type
XTH	Xyloglucan Endotransglycosylases/Hydrolases
XYL	Beta-D-Xylosidase

Scope and outline of the thesis

Growth is often exponential, whether it is a bacterial colony, a plant leaf or the human population. In the case of us, humans, the recent population growth is breaking all records and is causing an enormous pressure on the entire ecosystem. To safeguard the diversity of life on our planet, including ourselves, there is a great need for more sustainable consumption and production practices. One aspect of a more sustainable society is the use of less land for the same agricultural production. Drought is one of the major factors that reduce yields worldwide. When drought is severe it will cause wilting and eventually it will kill the plant, but when drought is rather mild it can still severely reduce growth and hence yields, without threatening the survival of the plant. This thesis specifically focuses on the effect of mild drought stress on the growth of leaves, the main energy producing organs of the plant.

To give the reader solid background knowledge in this specific matter, the **introduction** discusses the response of plants upon drought and what is already known on growth inhibition upon stress, with emphasis on leaf growth. The main part of this thesis describes the use of natural variation to unravel growth responses upon mild drought stress. Therefore, the second part of the introduction discusses the use of natural variation in studying the genetic architecture of complex traits in *Arabidopsis thaliana*.

In **Chapter 1**, 25 mutants that are tolerant against severe stress are screened in an automated way for enhanced growth in a controlled mild drought treatment. The outcome and implications for drought research are discussed.

In **Chapter 2**, the methodology is described that was used to investigate the responses of leaves to mild drought stress, from early in development until maturity. A set of six accessions was subjected to this set-up and different leaf growth-related phenotypes were measured, ranging from the macroscopic level, such as whole rosette area, to microscopic features, such as cell area. Furthermore, the transcriptome response of young developing leaves was characterized. Both phenotypic and transcriptional responses to mild drought stress are described.

In **Chapter 3**, different approaches were conducted to retrieve genes that are involved in the mild drought response. The phenotype and transcriptome responses and their variability over a large set of accessions are described. A modeling approach that delivered a set of genes that show a distinct expression response to the drought is discussed. Ultimately, two different GWAS approaches that associated phenotype and transcriptome data, respectively, with the genotypes of the different accessions are presented.

Finally, the **conclusions and future perspectives** give a concise overview of the main results of the different chapters and a discussion on improvements of the different analyses with future opportunities of studying natural variation to unravel the genetic architecture of complex traits.

Summary

Due to their sessile lifestyle, plants are limited in their possibilities by their close environment. The state of the plant's close environment will determine its morphology and life-cycle and water availability will be one of the key determinants. To cope with reduced water availability, plants have acquired numerous adaptations. One of these adaptations is the active growth reduction. Growth inhibition caused by drought - even when stress levels are mild - leads to major yield losses worldwide. However, the exact underlying regulatory mechanism is still poorly understood.

Because water is so important, its availability imposes great evolutionary pressure on plants. Therefore isolated populations that are located in different habitats will be adapted to the differences in water availability. The model plant *Arabidopsis thaliana* knows a global distribution, covering Eurasia, Japan, North America and even some patches along the African coastline. Because Arabidopsis is mainly selfing, the populations are isolated and have acquired different adaptations. Recent genotyping of different Arabidopsis accessions delivered the data that contains the blueprints of these adaptations. Methods like genome wide association mapping (GWAS) can be used to extract the relevant information for a specific phenotype form this genotypic data.

This thesis starts with screening the response to mild drought of Arabidopsis mutants that were described in literature as tolerant against severe stresses. However, none of the mutants performed better in mild drought stress, concluding that distinct mechanisms are involved in severe and mild drought stress. We set out to investigate the mechanisms involved in the mild drought response, using natural variation in Arabidopsis. In different accessions the response to mild drought stress of leaf growth-related phenotypes was measured, ranging from whole rosette area and specific leaf area measurements at different developmental stages, to detailed characterizations of cellular parameters such as number and area of epidermal cells and stomatal index. Furthermore, the transcriptome of young developing leaves was characterized to determine the genes that are transcriptionally affected by the mild drought stress in growing tissue.

A first exploration of the natural variation in six accessions showed that the different leaf-growth related phenotypes were affected by the treatment and that the effect differed between accessions. The transcriptome data allowed for identifying a set of 354 genes showing a common response to the treatment in the six accessions. Main mechanisms that were involved in this response were ABA, proline and cell wall modifications.

To be able to draw links between the phenotypic and genetic natural variation, the above listed phenotypes were characterized in a collection of 98 accessions and subsequently subjected to GWAS. The different phenotypes showed substantial variation between the accessions in their response to the mild drought stress. Some of these phenotypic differences associated with variations in the genotype. At the location of the associated genetic variants a number of genes (*mi*R*171c*, *CSLC4*, *SAP12* and *EXL1*) could be found with a plausible function in the mild drought response of the specific phenotypes.

In addition the phenotypic characterization also the transcriptional response upon mild drought of 89 accession was determined. Gene expression was highly variable in the different accessions. The gene expression information was used in a modeling approach that defined a set of 283 genes that can distinguish mild drought from control treated samples. While this set of 283 genes of stress predictors was defined without prior knowledge, it is indeed enriched for genes involved in drought responses and growth regulation. In order to shed light on the regulation of the transcriptional response upon mild drought stress, the transcriptome data was associated with the genotype in the expression GWAS (eGWAS). This showed that regulation of the differential expression was located in *trans* and covered different types of regulators such as transcription factors, a peptide hormone, epigenetic regulators and transposable elements.

This study is a first step in unraveling the natural variation of the mild drought response of growing tissue. The presented phenotypic, genotypic and transcript data can, to our opinion, be further exploited. By integrating results from different GWAS-approaches with additional co-expression analyses and large-scale literature screening it will be possible to determine the regulatory networks that underlie specific phenotypes. Constructing this regulatory network will allow for specific engineering of the network in order to breed crops for a more sustainable agriculture.

Samenvatting

Door hun sessiele levensstijl zijn planten beperkt in hun mogelijkheden tot wat de plaats waar zij kiemden te bieden heeft. De toestand waarin hun directe omgeving verkeerd is dan ook allesbepalend voor de vorm en leven van de plant, waarbij de beschikbaarheid van water één van de belangrijkste zal zijn. Omdat de beschikbaarheid van water variabel is, hebben planten tal van aanpassingen verworven die hen in staat stellen om het ook met minder water te doen. Eén van die mechanismen is een inhibitie van de groei. Het is die groei-inhibitie - zelfs al is de droogte vrij mild – die leidt tot enorme oogstverliezen in de landbouw. Het is tevens ook die groei-inhibitie die nog steeds amper gekend is.

Omdat water net zo belangrijk is voor planten, zal de beschikbaarheid ervan voor een grote evolutionaire druk zorgen. Hierdoor zullen geïsoleerde populaties die in verschillende habitatten leven zich hebben aangepast aan de verschillen in waterbeschikbaarheid. Het model organisme *Arabidopsis thaliana* is verspreid over een groot stuk van de wereld, met populaties in Eurazië en Japan tot Noord-Amerika en bepaalde kustgebieden in Afrika. Deze verschillende populaties zijn aangepast aan hun lokale omgeving en dus ook aan verschillen in de beschikbaarheid van water. De blauwdruk van deze aanpassingen ligt versleuteld op het DNA van de verschillende Arabidopsis accessies. De data met deze blauwdrukken werd recent aangereikt door de genotypering van een groot aantal van deze accessies. Methodes zoals 'genome wide association mapping' (GWAS) kunnen gebruikt worden om de relevante informatie voor een specifiek kenmerk van de plant te onttrekken.

Deze thesis start met een analyse van de respons onder milde droogte in mutanten waarvoor eerder een tolerantie tegen extreme stress werd geobserveerd. Gezien geen van de mutanten beter presteerde in milde droogte werd geconcludeerd dat er afzonderlijke mechanismen actief zijn in milde en extreme droogte. Een verdere zoektocht naar de mechanismen die actief zijn in milde droogte was dus nodig. Hiervoor werd de natuurlijke variatie in respons op milde droogte in Arabidopsis onderzocht. Voor de verschillende accessies werden in controle en milde droogte condities verschillende bladgroei-gerelateerde fenotypes gekarakteriseerd; variërend van globale rozet- en bladoppervlakte metingen tot karakterisering van specifieke cellulaire kenmerken zoals aantal en oppervlakte van epidermiscellen en stomatale index. Om een idee te krijgen van de genen die actief betrokken zijn in de reactie tegen milde droogte van groeiende bladeren werd ook het transcriptoom gekarakteriseerd van jonge, nog ontwikkelende bladeren.

Een eerste analyse van zes accessies toonde aan dat de verschillende fenotypes werden beïnvloed door milde droogte en dat de invloed verschillend was tussen accessies. Transcriptioneel was er een groep van 354 genen die een gelijkaardige respons vertoonden tegen de milde droogte in de zes accessies. Uit deze set van genen bleek een belangrijke rol weggelegd voor abscisinezuur, proline en cel wand modificaties.

Om een verband te vinden tussen de natuurlijke variatie in de fenotypes en de genetische verschillen werden de hierboven opgelijste fenotypes in een collectie van 98 accessies gekarakteriseerd en vervolgens onderworpen aan GWAS. De verschillende fenotypes toonden aanzienlijke verschillen tussen de accessies in hun respons tegen de milde droogte. Een aantal van deze fenotypische verschillen associeerden met variaties in het genotype. Op de plaats van die

geassocieerde genetische verschillen troffen we genen (*miR171c*, *CSLC4*, *SAP12* and *EXL1*) aan die zeer waarschijnlijk betrokken zijn in de respons tegen milde droogte van de fenotypes.

Naast de fenotypes werd ook de transcriptionele respons onder milde droogte gekarakteriseerd. De gen expressie data werd onder andere gebruikt bij het opstellen van een classificatiemodel dat resulteerde in een set van 283 genen die duidelijk het onderscheid konden maken tussen controle en milde droogte behandelde stalen. Deze set van 283 genen, geselecteerd zonder voorafgaande kennis, bevatte een significant aandeel aan genen die gekend zijn voor hun rol in droogte of regulatie van groei. Om meer te weten te komen over de regulatie van de transcriptionele respons in milde droogte werd de transcriptoom data geassocieerd met de genotypes van de verschillende accessies in een expressie GWAS (eGWAS). Daaruit bleek dat differentiële expressie in milde droogte vooral vanuit *trans* gelokaliseerde loci werd gereguleerd. De regulerende loci besloegen verschillende types regulatoren; transcriptie factoren, ene peptide hormoon, epigenetische regulatoren en transponeerbare elementen.

Deze thesis is een eerste stap in het ontrafelen van de natuurlijke variatie in de respons tegen milde droogte van groeiende bladeren. De fenotypische, genotypische en gen expressie data die hier beschreven zijn kunnen, volgens ons, nog verder geëxploreerd worden. Door de resultaten van verschillende types GWAS te integreren met co-expressie analyses en grootschalige literatuurstudies zal het mogelijk worden om de regulatorische netwerken van specifieke fenotypisch te gaan blootleggen. De constructie van zo'n regulatorisch netwerk zal toelaten om zeer gericht het netwerk te gaan aanpassen in functie van gewasontwikkeling voor een duurzame landbouw.

Introduction – Plant life in drought

From the moment the algal ancestor of land plants moved from the aquatic to the terrestrial environment, it got confronted to a reduced availability of water. In addition, since plants are sessile organisms, they are exposed to numerous fluctuations in the environment during their life cycle from which they cannot simply escape. During the course of evolution, a wide range of adaptations has allowed plants to successfully colonize different habitats, including extremely harsh conditions as can be found in deserts. However, drought stress can lead to decreased plant performance and since plants are primary producers in numerous ecosystems, this abiotic stress often has an impact on the higher trophic levels of the food chain.

Human society is largely relying on plants and plant-derived products. Numerous plant species constitute a big part of our daily diet directly as fruit, grains, nuts and vegetables, and indirectly as feed for livestock. Besides the use of plants for nutritional purposes, crops such as cotton, flax and hemp are grown for fiber and many crops are grown to extract sugar or oil to be transformed into biofuels. Growing all those plants requires large amounts of water. Agriculture is currently responsible for 70% of the worldwide human water consumption (UNESCO, 2001). Water is essential for agriculture and with over 40% of the crops being produced under irrigated conditions (Döll and Siebert, 2002), stability in food production is very sensitive to changes in water availability. Droughts are however expected to increase in frequency, intensity and duration due to climate change (IPCC, 2014). Recent drought episodes have shown that the reduced agricultural production may cause local food crises and famine in developing countries (FAO, 2011) and can increase food prices globally (The World Bank, 2012). Studying the adaptations and responses of plants to drought is thus not only interesting from an academic point of view, but will also deliver solutions for a more efficient crop production which can withstand more easily the ever-fluctuating environment.

A. Responses to drought

Plants perceive drought as a reduction in water availability in the soil. In this sense, drought stress is similar to other abiotic stresses that reduce the availability of water such as osmotic, salt and freezing stress, but drought stress typically refers to the situation in soil. However, drought stress will in the field often combine with high temperature, high light and increased mineral concentrations causing salinity stress. When subjected to drought stress, a plant may respond at various organizational levels, from molecular to morphological changes, and even changes in the life cycle duration. The different physiological responses of plants to resist drought conditions can be divided into two contrasting categories: stress avoidance and stress tolerance (Figure 1; Verslues et al., 2006).

A.1 From stress avoidance to stress tolerance

In first instance, when a reduction in water availability is perceived, plants attempt to avoid the stress. To do so, they maintain homeostasis by readjusting the internal water potential by decreasing water loss or by increasing the water uptake. Water loss takes place at the leaf surface

where a part will evaporate through the cuticle, but the major amount of water is lost via the stomata. These small structures, spread over the surface of the leaf, are the portals between the plant and its surrounding atmosphere. Stomata allow CO_2 to enter the plant for assimilation into sugars and simultaneously regulate the evaporation of water in order to control leaf temperature. Stomatal closure is one of the first responses of the plant upon drought stress and is an effective way to reduce water loss on a short term. On the longer term, the cuticle will become thicker due to an increased synthesis of waxy esters (Kosma et al., 2009) allowing a decrease in water evaporation but also an increase of sunlight reflection to maintain a low leaf temperature.

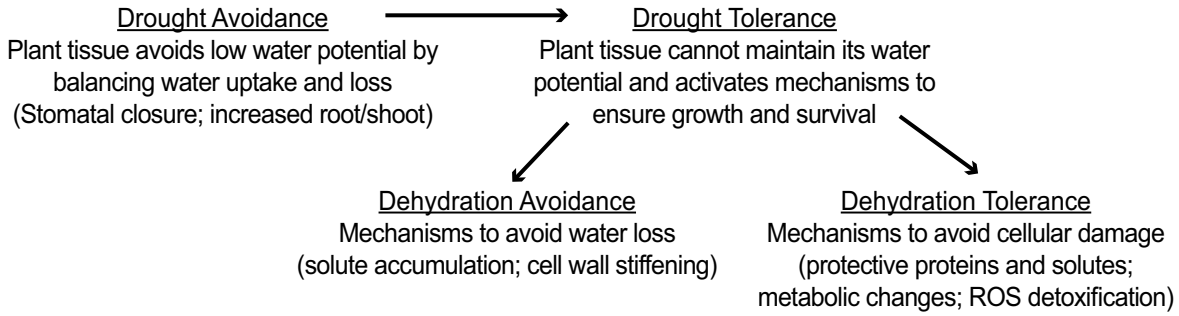


Figure 1: Drought tolerance/avoidance model. When the stress is not severe, the drought avoidance mechanisms will balance water uptake and loss. If the drought persists or increases in severity, the drought tolerance mechanisms get activated in order to ensure growth and survival of the plant. Dehydration avoidance mechanisms will try to maintain the water content. Whereas dehydration tolerance mechanisms are avoiding cellular damage if the water content cannot be maintained (Verslues et al., 2006).

To increase the uptake of water under drought, root growth is stimulated to allow the plant to drain water from new sources. However, if the water potential of the plant and the soil become equal, the plant will need to decrease its water potential accordingly to assure the water uptake. Plant water potential is equal to the sum of the osmotic potential and the turgor pressure, both regulated by different means (Verslues et al., 2006). The osmotic potential can be lowered by accumulating different compatible solutes in a process known as the osmotic adjustment (Zhang et al., 1999), while the turgor pressure is adjusted by affecting the elasticity of the cell wall.

Well known compatible solutes are proline and glycine betaine, that both can accumulate to high levels without interfering with cellular processes (Yancey et al., 1982). Besides their function in restoring plant water potential, compatible solutes may also have protective and signaling functions. Proline and glycine betaine maintain protein folding and can stabilize the redox balance, which is affected by ROS production under drought. Accumulation of proline and glycine betaine also induces the expression of several transcription factors, ROS-scavengers and low-temperature-responsive genes that can mediate the stress response (Szabados et al., 2011). However, the drawback of accumulating these compatible solutes is that their production is often energy and resource intensive (Krämer and Boyer, 1995).

The turgor pressure of the cell can be modified by adjusting the elasticity of the cell wall (Krämer and Boyer, 1995; Murphy and Ortega, 1995). That is because the elasticity corresponds to the pressure that is needed to make a change in cell volume. So if the cell wall is rigid, more pressure is required to increase the cell volume than when the cell wall is elastic. But this also works in the other direction, so it will require more pressure, or say energy, to decrease the cell volume by losing water. In other words, it will require more energy to lose water from the cell if the cell wall is rigid. Cell wall stiffening is therefore a widely used strategy in non-growing tissue to maintain the water potential of the plant.

With increasing severity of the drought, maintaining the water potential and avoiding cellular damage becomes more difficult. Therefore plants need to cope with the reduced water content by inducing dehydration tolerance mechanisms (Verslues et al., 2006). Examples of plants tolerating periods of extreme drought are extremophiles such as resurrection plants; e.g. *Craterostigma plantagineum*, *Craterostigma wilmsii*, *Haberlea rhodopensis*, *Sporobolus stapfianus*, *Tortula ruralis* and *Xerophyta humilis* (Gechev et al., 2012; Griffiths et al., 2014). These plants can survive months or even years of drought in a metabolic dormant state and are using the same mechanisms present in non-extremophile plants to cope with drought. To protect the cellular structures against extremely reduced water availability, protective proteins such as the late embryogenesis abundant (LEA) proteins accumulate. The Arabidopsis genome contains 51 LEA encoding genes, which are mostly expressed in seeds and vegetative tissue exposed to dehydration stresses such as cold, drought and osmotic stress. The expression of LEAs under stress is mainly regulated by abscisic acid (ABA), through the presence of ABA responsive elements (ABREs) and drought responsive/C-repeat/low-temperature responsive elements (DR/CRT/LTREs) in the promoter regions of LEAs. Several LEAs have a role in stabilizing labile enzymes under dehydration but their high sequence diversity suggests various modes of action for the yet functionally poorly characterized LEAs (Hundertmark and Hincha, 2008).

Drought stress will, similar to other biotic and abiotic stresses, induce the production of reactive oxygen species (ROS). Because of their highly reactive nature, ROS are harmful for all processes and structural components of the cell but they also have a signaling role in various stresses. Therefore a strict balance needs to be maintained between ROS production and scavenging. To prevent cellular damage, ROS scavengers such as superoxide dismutase, catalase, ascorbate peroxidase, glutathione reductase and dehydroascorbate reductase are produced in stress conditions as part of the stress tolerance responses (Dat et al., 2000; Mittler et al., 2011).

The above-described mechanisms allow plants to tolerate a decrease in water availability but do not necessarily cause a more efficient water usage. The amount of carbon that is assimilated per water molecule is described as the water use efficiency (WUE) and depends on photosynthetic efficiency, the amount of water lost through evapotranspiration and the efficiency of water uptake via the roots. The main determinant of the efficiency of photosynthesis, under optimal conditions, is the RUBISCO enzyme (Stitt and Schulze, 1994). Attempts to improve the kinetics of RUBISCO were, however, unsuccessful (Parry et al., 2013; Whitney et al., 2011). Evapotranspiration is taking place at the leaf surface and mainly through the stomata. Stomatal closure and reduced permeability of the leaf surface due to a thick cuticula prevents water loss. On the downside, restricting evapotranspiration also reduces the uptake of CO_2, causing a trade-off between water loss and carbon uptake. In order to increase the water uptake a greater and deeper root system may reach more or richer sources of water, but also a lower water potential in the roots may allow to extract more water from the soil. The combination of sufficient water uptake, low evapotranspiration and an efficient photosynthetic system will result in an efficient water usage.

In conclusion, plants possess different mechanisms, functioning from cell to whole plant level to cope with drought stress. The activation of one or the other response depends on the severity of the stress and whether the plant can maintain its water potential or not.

A.2 Abscisic acid (ABA): A main regulator of drought responses

Many of the drought responses, such as stomatal closure or LEA accumulation, are regulated by the phytohormone ABA known as the main drought response-regulating hormone. Initially being described as a growth inhibitor of cotton fruit ("abscisin II"; Ohkuma et al., 1963), ABA is also involved in embryo maturation, seed dormancy, germination, cell division and elongation, floral induction, and responses to drought, cold, salinity, pathogens and UV radiation (Finkelstein, 2013). Numerous studies on ABA signaling have resulted in the identification of a "core ABA signaling pathway" (reviewed by Cutler et al., 2010).

Upon drought stress, ABA synthesis is initially induced in the roots and as soil moisture further decreases, ABA also gets produced in the leaves (Bahrun et al., 2002). The signaling machinery is activated once the presence of ABA is detected by the receptors (Figure:2), soluble proteins members of the family of pyrabactin resistant/PYR-like/regulatory component of ABA receptor (PYR/PYL/RCAR) proteins (Park et al., 2009; Miyakawa et al., 2013). In Arabidopsis, the PYR/PYL/RCAR family contains 13 functional ABA receptors (PYR1 and PYL1-PYL12) (Fujii et al., 2009). Once ABA is bound to one of the receptors, the conformation of this protein will change, allowing for stable interaction with members of the protein phosphatases 2Cs (PP2Cs). PP2Cs are Ser/Thr protein phosphatases (Leung et al., 1994; Meyer et al., 1994; Leung et al., 1997; Rodriguez et al., 1998) that negatively regulate the ABA response (Sheen, 1998; Rubio et al., 2009) by inhibiting the activity of SNF1-related kinases (SnRKs). Binding of PYR/PYL/RCAR-ABA complex to the PP2Cs impedes the inhibitory effect of the PP2Cs on the SnRKs, allowing them to activate transcription factors, ion channels and other mediators of the ABA response. Transcription factors involved in the ABA response mainly bind ABA response elements (ABREs), "coupling element" (CE1) and CE1-like or RY/Sph elements present in the ABA regulated gene promoters (reviewed by Finkelstein, 2013).

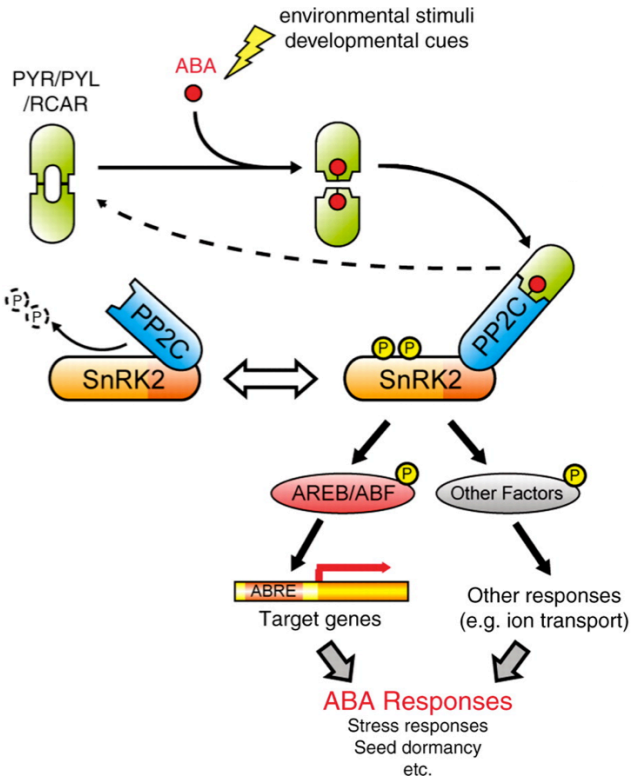


Figure:2: ABA signaling cascade. Model of the major ABA signaling pathway as proposed by Umezawa et al. (2010). PYR/PYL/RCAR, PP2C and SnRK2 proteins form a signaling complex referred to as the 'ABA signalosome'. Under normal conditions, PP2C negatively regulates SnRK2 by direct interaction and dephosphorylation of multiple residues of SnRK2. Once abiotic stresses or developmental cues up-regulate endogenous ABA, PYR/PYL/RCAR binds ABA and interacts with PP2C to inhibit protein phosphatase activity. In turn, SnRK2 is released from PP2C-dependent regulation and activated to phosphorylate downstream factors, such as the AREB/ABF bZIP-type transcription factors or membrane proteins involving ion channels. Figure derived from Umezawa et al., 2010.

In total 38 SnRKs have been identified in Arabidopsis that, besides their response to ABA, are strongly suggested to be important convergence points for different metabolic, hormonal and stress signals during growth and development (Rodrigues et al., 2013). Besides SnRKs, the PP2Cs may also dephosporylate other kinases such as the calcium dependent protein kinases (CPKs) and some of the ABA induced transcription factors and ion channels. Through these dephosphorylations the PP2Cs may further inhibit the ABA response or restore homeostasis (reviewed by Cutler et al., 2010; Finkelstein, 2013).

Extensive studies of various stresses and ABA treatments reveal two waves of transcriptional responses: an early transient response around 3hrs after stress initiation and a second, sustained response from ±10hrs onward (reviewed in Finkelstein, 2013). The early response mainly induces regulatory proteins that can control the ABA response such as transcription factors, kinases, phosphatases and the so-called 'early response to dehydration' (ERD) genes (Yamaguchi-Shinozaki and Shinozaki, 2006; Fujita et al., 2011). The ERD genes encode a functionally heterogeneous group of proteins, containing a chloroplast ATP-dependent protease (ERD1), a heat shock protein (ERD2, ERD8), a membrane protein (ERD4), a proline dehydrogenase (ERD5), a carbohydrate transporter (ERD6), a ubiquitin extension protein (ERD16), glutathione-S-transferases (ERD9, ERD11, ERD13), LEAs (ERD10, ERD14), a protein interacting with poly-A-binding proteins (ERD15) and proteins with yet unknown functions (ERD3, ERD7) (Kiyosue et al., 1993; Kiyosue et al., 1994a; Kiyosue et al., 1994b; Kiyosue et al., 1998; Nakashima et al., 1998; Kariola et al., 2006; Mitra et al., 2007; Soitamo et al., 2008; Alves et al., 2011). The ERD genes are differentially expressed upon drought stress and/or ABA treatment and have important functions in the drought response e.g. ERD5 that catabolizes proline, an important osmolyte, is down-regulated upon drought (Kiyosue et al., 1996). ERD15, a negative regulator of the ABA response, improves drought tolerance when silenced (Kariola et al., 2006).

The late responsive genes generally encode proteins with functions similar to LEAs, proteases, chaperones, proteins involved in compatible solute metabolism, reactive oxygen (ROS) detoxifiers and ion- and water-channel proteins such as aquaporins (Ingram and Bartels, 1996). Aquaporins are transmembrane proteins, localized at the tonoplast (Tonoplast Intrinsic Proteins, TIPs) or at the plasma membrane (Plasma membrane Intrinsic Proteins, PIPs), that control the water flow across cell membranes. Upon drought stress, ABA has been shown to down-regulate the aquaporin expression in bundle sheath cells (Shatil-Cohen et al., 2011) and to inactivate aquaporins through dephosphorylation (Kline et al., 2010), leading to a reduced water flux. However, in maize plants exposed to osmotic stress, ABA was found to induce aquaporin expression, especially in roots. The resulting increase in hydraulic conductivity may promote water uptake by the roots and accelerate the recovery of the leaf water potential (Parent et al., 2009). The regulation of aquaporins by ABA is clearly playing a role in the drought response of the plant, yet the exact function still needs to be further elucidated. It is likely that an increase or decrease of aquaporin expression will depend on the plant organs and the severity of the stress.

A.3 ABA-independent regulation of drought responses

Besides the well-described ABA-regulation of drought responses, an ABA-independent transcriptional regulation has been reported. The promoter regions of the ABA-independently regulated genes contain drought responsive elements (DRE), the similar C-repeat (CRT) or low

temperature responsive element (LTRE). Eight transcription factors in Arabidopsis can bind the DRE/CRT/LTRE domains; DREB1A/CBF3, DREB1B/CBF1, DREB1C/CBF2, DREB1D/CBF4, DREB1E/DDF2, DREB1F/DDF1, DREB2A and DREB2B (Stockinger et al., 1997; Liu et al., 1998; Sakuma et al., 2002). The DREB1 transcription factors are mainly induced by cold, except for DREB1D/CBF4, which is induced by osmotic stress. The DREB2 proteins are transcription factors that function in the dehydration and salinity stress response (Nakashima et al., 2000; Sakuma et al., 2002). DREB2A requires post-translational activation and when constitutively expressed in its active form leads to growth retardation (Liu et al., 1998). ABA-dependent and ABA-independent regulation of transcription is however not mutually exclusive as exemplified by *RD29A*, a cold and drought inducible gene that is regulated via both pathways (Narusaka et al., 2003).

B. Mild versus severe drought

Our knowledge on how plants respond to drought is mainly obtained from experiments in which plants were subjected to severe drought or osmotic stress. Many of the mechanisms described above were discovered by performing survival assays or exposing plants to extremely high concentrations of osmotica, threatening the survival of the plants. Mild drought on the other hand, does not induce wilting, chlorosis, necrosis or other signs that suggest that the plant's life is in danger. In this thesis we define mild drought as a stress level that allows the survival of the plant and that triggers a clear growth penalty (±60% reduction in leaf area). Severe and mild drought stress activate different mechanisms and mutants with increased tolerance against severe stresses are not per se better growing under mild drought stress (Skirycz et al., 2011b). Studies on mild drought stress response, however, do report clear phenotypic and transcriptional changes (Aguirrezabal et al., 2006; Pereyra-Irujo et al., 2008; Harb et al., 2010; Baerenfaller et al., 2012; Marais et al., 2012; Clauw et al., 2015), but these responses are different from the observations in severe drought. In general, plants exposed to mild drought are not threatened in their survival and are therefore not showing extreme stress phenotypes such as bleaching or altered germination rates (reviewed by Claeys et al., 2014b), but they clearly reduce their growth. Transcriptionally the response to mild and severe drought also differs, which results in different marker genes for either stress severity. Good examples are ERF5, MYB51 and WRKY33 which are induced at low concentrations of the osmoticum mannitol and are good marker genes for this mild osmotic stress (Skirycz et al., 2010; Skirycz et al., 2011a). Different dosages of osmotic, salt and oxidative stress ranging from very low (e.g. 5mM mannitol) to extreme stress levels (e.g. 300mM mannitol) resulted in different alterations of growth and transcriptome, showing that stress responses are dosage dependent (Claeys et al., 2014b).

Since the severe drought responses are relatively well described in literature and as mild stress responses seem to clearly differ we opted in this thesis to unravel the mild drought response. Moreover, as discussed above, severe drought will completely stop plant growth while in rather mild drought plants keep growing at an adapted level. It is this balancing regulation of growth that intrigued us from an academic point of view. But also in terms of agricultural applications, it is interesting to find out the mechanisms that are altering the growth without completely stopping it, in order to obtain the knowledge to fine-tune growth responses to drought in crops.

For this research, the leaf was chosen as a model organ since they are more susceptible to growth reductions upon mild drought than roots. Also leaf growth is of great importance since leaves produce the energy for the entire plant. Therefore the subsequent part of the introduction will present the current knowledge on the growth response of leaves exposed to drought.

C. Leaf growth responses to drought

Plants subjected to drought rapidly inhibit their leaf growth, in a so-called acute response, once the drought is stable or diminishes growth can recover or acclimate to the new environmental conditions, this is the adaptation response (Figure 3; Skirycz and Inzé, 2010). The acute and adaptation responses are nicely exemplified in barley leaves, which were shown to almost completely stop their growth within seconds upon salt stress, but resume their growth at adjusted growth rates about twenty to thirty minutes after the start of the salt treatment (Fricke, 2006).

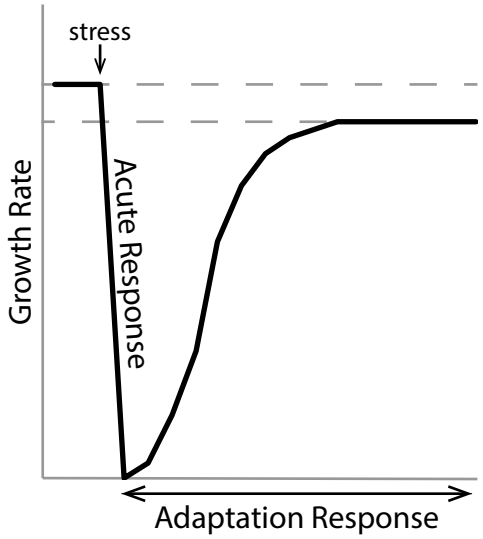


Figure 3: Acute and adaptation response upon stress. Growth rates will quickly decline when stress (e.g. drought) is sensed. Once the stress severity stabilizes, the growth rates adapt to the new situation during the adaptation response.

Growth is driven by two main processes, cell division and cell expansion, which are both affected by drought (Aguirrezabal et al., 2006; Baerenfaller et al., 2012; Clauw et al., 2015). During leaf development under standard conditions the leaf initiates as a cluster of cells emerging from the shoot apical meristem. These cells, that constitute the leaf primordium, will divide during the so-called proliferation phase. After a few days of cell proliferation, a cell cycle arrest front initiates at the tip of the leaf. Cells will then stop dividing and start differentiating and expanding. During the transition phase the cell cycle arrest front moves towards the base of the leaf until all cells are differentiating and expanding. Once all cells expanded, the leaf reaches its final size and is mature (Gonzalez et al., 2012). Because cell expansion takes place, to some extent, once cell division has ceased, drought mediated reductions in cell division can be compensated for by an increased cell expansion afterwards. Yet, also cell division can compensate for reductions in cell expansion (Skirycz et al., 2011a). Since most cells in the leaf stopped dividing by the time that most cells are expanding, such a compensation mechanism requires cells that retained their capacity to divide. Meristemoids are cells that are capable of dividing once cell proliferation has ceased. These cells

are spread out over the entire leaf epidermis and give rise to stomatal guard cells (Peterson et al., 2010; Pillitteri and Dong, 2013). Besides guard cells, meristemoids also form pavement cells. Up to 48% of the pavement cells in the first leaf may originate from meristemoids (Geisler et al., 2000; Bergmann and Sack, 2007), this mechanism thus potentially has a large effect on final cell number. Because of the potentially great contribution to total cell number, meristemoids are good candidates for compensating for reductions in cell number that happened earlier in leaf development.

The growth reduction under drought was long thought to be a secondary effect of the limited carbon availability due to stomatal closure and the reduced availability of water. However, drought induced growth inhibition is even observed when the turgor pressure is still sufficient to allow expansion (Michelena and Boyer, 1982). Likewise, the carbon balance was found to be positive under drought (Muller et al., 2011) and under osmotic stress an accumulation of starch has been observed (Skirycz et al., 2010). The observations that plants still reduce growth despite a positive carbon balance and adequate water potential lead to the hypothesis that the growth reduction is actively regulated.

D. Regulatory mechanisms of growth reduction in drought stress

The underlying mechanisms of how drought is affecting growth are still poorly understood but several pathways have been elucidated in the past decade (reviewed by Claeys and Inzé, 2013). As discussed above, drought affects both cell proliferation and cell expansion in developing leaves.

Cell proliferation is mainly regulated by controlling the transition between the different cell cycle phases by the cyclin-dependent kinases (CDKs) and their interacting cyclins (De Veylder et al., 2007). Three different mechanisms control the activity of the CDK-cyclin complexes (Fig. 4): the cyclin protein levels can be controlled through degradation induced by complexes such as the ANAPHASE-PROMOTING COMPLEX/CYCLOSOME (APC/C); phosphorylation of the CDKs can cause activation or inhibition of the CDK-cyclin complexes; the complexes can be inhibited by the binding of two main families of inhibitory proteins, the CYCLIN-DEPENDENT KINASE INHIBITOR (CKI)/KIP-RELATED PROTEIN (KRP)-type proteins and SIAMESE/SIAMESE-RELATED (SIM/SMR)-type proteins (Komaki and Sugimoto, 2012). All three CDK-cyclin regulating mechanisms are affected by drought and other abiotic stresses and can, in response to the stress, regulate the reduction in cell proliferation. An inhibitory effect of stress on CDKA activity itself has been described for mild osmotic stress in wheat (Schuppler et al., 1998) and Arabidopsis (Skirycz et al., 2011a), and for mild drought stress in maize (Granier et al., 2000).

Besides the direct regulation of the activity of CDK and CDK-cyclin complexes, a number of transcriptional responses upon stress can affect cell division. The expression of numerous cyclins is reduced by salt stress (Burssens et al., 2000), repressors of the APC/C complex such as DP-E2F-LIKE1 (DEL1) and UV-B-INSENSITIVE (UVI4) are down-regulated under mild osmotic stress (Claeys et al., 2012; Claeys et al., 2014a) and expression of CKI/KRP- and SIM-type CDK inhibitors is induced (Pettkó-Szandtner et al., 2006; Peres et al., 2007).

Also several phytohormones such as ethylene, gibberellic acid (GA), ABA and auxin play a role in controlling the growth response in drought and osmotic stress (Fig. 4). Studies on the osmotic stress response have attributed an important role for ethylene in regulating growth responses upon mild stress. In plants grown on low concentrations of mannitol (25mM), 1-aminocyclopropane-1-carboxylic acid (ACC) accumulates in growing leaves where it is converted to ethylene. Ethylene activates a signalling cascade involving MAP KINASE3 (MPK3) and MPK6, which activates ETHYLENE RESPONSE FACTOR5 (ERF5) and ERF6 by phosphorylation. The active ERF5 and ERF6 activate on one hand a stress responsive transcriptional cascade through direct induction of *WRKY33*, *STZ* and *MYB51*, on the other hand both ERFs inhibit growth by triggering *GA2*-OXIDASE6 (GA2-OX6) expression. The increased expression of GA2-OX6 leads to GA inactivation and reduced GA levels, which in turn stabilizes the DELLA proteins and ultimately inhibits leaf growth (Achard et al., 2009; Skirycz et al., 2010; Skirycz et al., 2011a; Dubois et al., 2013).

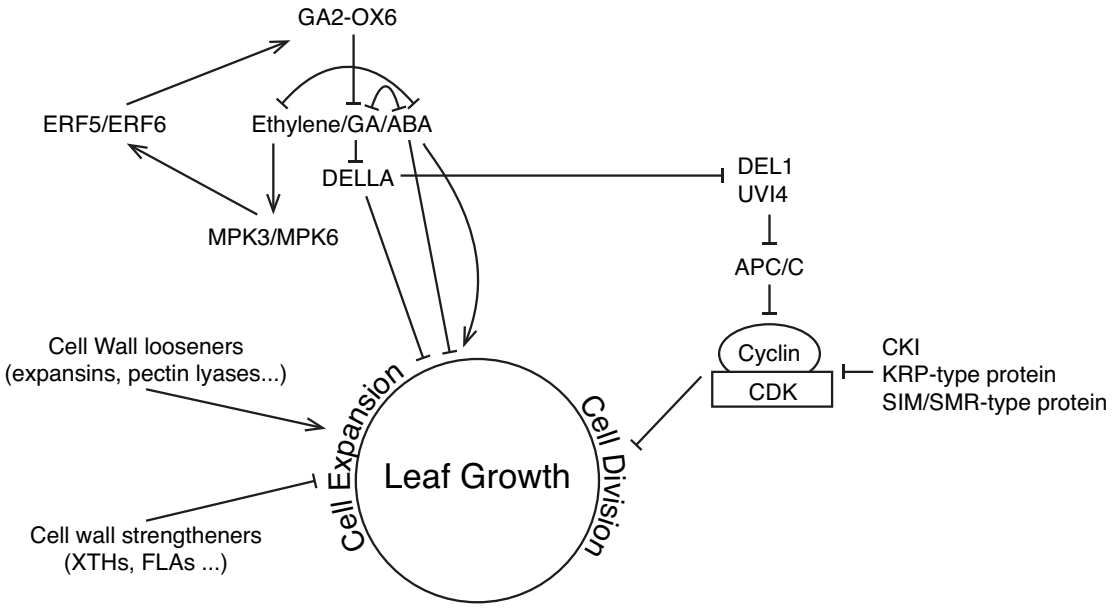


Figure 4: Overview of currently known pathways that regulate leaf growth upon mild drought stress. The CKI, KRP-type protein and SIM/SMR-type, DEL1 and UVI4 pathways are known to specifically affect cell division through interaction with the cyclin-CDK complex. While the cell wall modifying proteins are affecting cell expansion. Different plant hormones (ethylene, GA and ABA) are affecting leaf growth in general in an antagonistic manner.

The main drought response hormone, ABA, does not have one straightforward role in regulating growth. By affecting stomatal closure and maintaining water conductivity via aquaporins, ABA maintains the water potential and allows for growth through cell expansion. On the longer term, however, the decreased availability of CO_2, due to stomatal closure, will have a negative impact on biomass accumulation. ABA can also promote growth by reducing ethylene biosynthesis. Nevertheless increased endogenous ABA concentrations were also found to inhibit leaf growth (Tardieu et al., 2010; Wilkinson and Davies, 2010). Since ABA affects growth in different manners, the net growth reduction will depend on the combination of the different ABA effects. In addition, auxin was found to play a role in cell expansion responses upon salt and osmotic stress (Veselov et al., 2002; Veselov et al., 2008; Skirycz et al., 2010).

Cell expansion is to a large extent determined by cell wall modifiers (Fig. 4). In growing tissue exposed to drought, cell wall extensibility will be increased through the up-regulation of cell wall loosening expansins and pectin lyases, while the expression of cell-wall hardening enzymes such as xyloglucan endotransglycosylases/hydrolases (XTHs) is down-regulated. The increase in cell wall extensibility of growing tissues allows for maintaining the growth at lower turgor. This is in contrast with non-growing tissue, where the cell wall becomes more rigid upon drought in order to maintain the water potential of the tissue (Moore et al., 2008; Skirycz et al., 2010; Clauw et al., 2015).

Since water availability is such an important determinant for the plant's life and morphology, research has focused already for decades on drought responses. Many aspects of this response are well described. However, a high number of mechanisms are involved and show complex interactions. Being aware of the discrepancy between survival and growth, more research starts focusing on milder drought assays to unravel the regulatory network of the growth response to drought. Since drought is highly dosage dependent, many subtle players with small effects in this network might be at work. Therefore the question comes whether the current mutant studies offer the best tool to study subtle transcriptional changes shaping the complex growth response to drought. In the subsequent part of the introduction we will discuss the use of natural variation to identify the genetic architecture of complex traits, with a special focus on the model organism Arabidopsis.

Introduction – Natural variation in *Arabidopsis thaliana*

A. *Arabidopsis thaliana* around the world

Arabidopsis thaliana is a commonly found weed of the Brassicaceae family. Its small size (±30cm), small genome (±135Mb) and fast generation time (±6 weeks) makes Arabidopsis a model system for molecular, physiological and genetic research. The genome sequence of the Arabidopsis accession Columbia (Col-0) is available since 2000 (Arabidopsis Genome Initiative, 2000).

Figure 5 : Global distribution of *Arabidopsis thaliana*. Areas colored in red correspond to the continuous distribution of *A. thaliana*, red circles indicate additional sites. Figure derived from Krämer, 2015.

Arabidopsis is found in a wide range of habitats spread over the Eurasian continent, North-America, North-Africa and some regions along the African coast (Figure 5 ; Hoffmann, 2002). This broad geographic distribution exposed the different local populations to different evolutionary pressures, resulting in genetic adaptations to specific environmental conditions (Alonso-Blanco and Koornneef, 2000). These genetic differences in the numerous Arabidopsis accessions represent a highly interesting pool of information since they can indicate the genetic mechanisms underlying the adaptation to a specific condition. In contrast to mutant studies, in which the effect of perturbing one gene, often through a constitutive overexpression or knock-down, is studied, the natural variants are genetically differing in a more subtle and functional manner. Moreover, the difference observed between natural variants is not limited to one gene but entire regulatory networks are adapted to specific environmental conditions. Highly complex

traits will therefore show genetic adaptations on a large number of loci, each individually manifesting a small effect on the phenotype. In such cases, the individual loci may be difficult to detect in complex traits. Recently over 1100 accessions have been re-sequenced (www.1001genomes.org). This sequencing data allows for conducting association studies (Atwell et al., 2010), in which the genetic adaptations that correspond to certain phenotypic differences are determined. In this way the genetic architecture of complex traits can be unraveled.

B. Genetic diversity in *Arabidopsis thaliana*

A large part of the studies in Arabidopsis are conducted on the Col-0 reference accessions. However, this accession is not representative for the entire species. By sequencing the genomes of 18 accessions, 497,688 to 789,187 single-base differences, depending on the accession, were found with the Col-0 genome (Gan et al., 2011). The sequence differences were enriched for transposable elements and intergenic regions. About 17% of the bases deleted in one or more accessions were annotated as genes in Col-0 (Gan et al., 2011). Also between different accessions, substantial genetic variation is present. Whole genome sequencing of 80 accessions revealed more than 4 million single nucleotide polymorphisms (SNPs) and over 800.000 insertions and deletions (indels) of 1 to 20 bp (Cao et al., 2011).

The genetic differences between accessions are not random, some pairs of accessions are less different than others. How well accessions are genetically related is determined by the population structure. The population structure of the genetic polymorphisms was initially characterized by Sanger sequencing 876 short (500bp) dispersed genomic regions in 96 accessions. This showed that there is significant population structure among the Arabidopsis accessions. The different populations are genetically isolated by geographic distance, but the spreading is rather continuous than discrete (Platt et al., 2010). Moreover, a large fraction of the genetic polymorphisms are shared worldwide (Nordborg et al., 2005), showing that despite the high degree of selfing in Arabidopsis, outcrossing is still happening. The continuous spreading of the populations together with the large degree of shared polymorphisms makes that the differences between closely related accessions are too vague to create statistical sound phylogenetic trees to visualize the population structure.

Besides a first description of the population structure of Arabidopsis, Nordborg et al. (2005) also provided insight into the linkage disequilibrium (LD) pattern among the SNPs. LD measures the degree of non-random associations between alleles in a population and reflects the evolutionary history of recombination and outcrossing. The LD causes the sequence variants to be found in haplotype blocks containing one or more causal polymorphisms and a variable number of linked non-causal polymorphisms (Weigel, 2012). By analyzing 314,602 non-singleton SNPs in 19 accessions, LD was shown to decay within 10kb in Arabidopsis (Kim et al., 2007). The estimated LD in Arabidopsis is thereby sufficiently low to legitimate association mapping.

C. Studying natural variation: QTL mapping and GWAS

Arabidopsis is an excellent model species for association studies. Its extensive genetic variation and the low degree of LD allows for mapping genetic loci at a high resolution. Moreover, the selfing nature of Arabidopsis resulted in a high degree of homozygosity. The different populations therefore represent inbred lines that need to be genotyped only once but can be phenotyped over several generations.

Association studies, where the genotypic variation is linked to phenotypic variation, will preferentially follow three steps. First the phenotypic variation of the trait of interest is measured in different genetic backgrounds. Subsequently the genetic basis of the phenotypic variation is identified. Finally the molecular function of the different alleles can be unraveled (Koornneef et al., 2004).

C.1 Determining the phenotypic variation between accessions

Characterizing the phenotypic variation among different accessions gives a first impression of the potential genetic variation present in the collection under study. The phenotypes of the different natural variants, adapted to different environments, will therefore be determined in environmental conditions that are equal for the different accessions, a so-called common garden experiment. The trait of interest, will be characterized in a standardized manner and since a high number of accessions (min. ±100; Atwell et al., 2010) is required to successfully conduct an association study, researchers are often employing high-throughput phenotyping systems (Dhondt et al., 2013).

In theory, any trait can be subjected to an association study, as long as there is sufficient variation in the phenotype and the underlying genetic loci. Differences in phenotypes and the causative genetic loci are mainly caused by natural selection. Therefore, traits that are prone to natural selection are favorable to conduct association studies (e.g. flowering time, stress resistance, nutrient efficiency, etc.). Different traits may severely vary in the complexity of their genetic architecture. Traits like pathogen resistance can be determined, in the simplest case, by one gene encoding a compound that is toxic to the pathogen. The one genetic locus that is responsible for the variation in the functionality of the toxin will have a large phenotypic effect (resistance or not against the pathogen). While more complex traits, like organ growth, will be influenced by a large number of genes. All genes together can then explain the entire complex trait. Each of the genes separately will, however, explain only a small percentage of the trait. Therefore, it is worthwhile to split up a complex trait in its underlying "subtraits". For example growth, the underlying cellular parameters like cell division and cell expansion can be characterized as more simple traits that underlie the complex trait, growth. This approach has been used for root growth, where the underlying cellular traits (such as length of meristem zone, length of elongation zone and length of mature cortical cell) were successfully subjected to GWAS (Meijón et al., 2014). Since the genetic variation evolutionary mainly arose through natural selection in response to varying environmental conditions, natural accessions are ideally suited for characterization of the phenotypes in opposing treatments (such as biotic, drought, salt and light stress), as these can reveal the underlying genetic adaptations.

C.2 Associating genotype to phenotype

Once the phenotypic variation is characterized, the second step is to retrieve the loci that are responsible for the phenotype of interest. The phenotypic variation can be qualitative or quantitative. In the simplest case, the studied trait is qualitative and the variation in one single segregating locus is responsible for the observed phenotypic variation; the trait is therefore monogenic. In this case, the locus can be determined by standard Mendelian linkage mapping, where the position of the underlying gene can be mapped according the recombination frequency with a marker gene. Nevertheless, many, if not most, traits are quantitative in nature and will be shaped by a great number of small-effect loci. In this case the trait is multigenic. An example is plant size, where the multigenic character makes that many different allelic combinations define many different phenotypic outcomes. The large number of varying phenotypic outcomes of quantitative traits results typically in a normal distribution of the phenotype in a population. Fischer defined that the effects of the involved genes deviate the phenotype from a central value, the mean (Fischer, 1918). This definition allowed for the statistical analysis of quantitative traits and the underlying genetics. The advent of high-throughput genotyping methods and, recently, next generation sequencing techniques made it possible to determine the genetic variants in a population showing differences in a quantitative trait. Furthermore, analyzing methods such as quantitative trait locus (QTL) mapping and genome wide association mapping (GWAS) allow for associating specific genetic loci to the quantitative trait of interest. QTL-mapping will be mainly used to unravel the genetic differences between contrasting phenotypes. GWAS, on the other hand, can more be seen as a screen of the global genetic diversity underlying a specific trait. Both techniques are discussed and compared in the following part.

C.3 QTL-mapping

To perform QTL mapping, a mapping population needs to be created. Mostly the population will be formed by crossing two parental lines that show opposite phenotypes for the trait of interest. The obtained F1 generation is then crossed so that the parental alleles recombine (Figure 6A). Subsequently, the F2 is selfed for six or more generations to obtain homozygosity, finally resulting in a so-called recombinant inbred line (RIL) population. The advantage of these populations is that they only need to be genotyped once and can be phenotyped from generation to generation. Once the genotype and phenotype are determined for the different RILs, statistical approaches allow for associating genetic markers to the trait of interest. The accuracy of the QTL mapping is dependent on the segregation between the genetic markers and thus the number of recombination events. In Arabidopsis typical QTL regions cover 1.2 to 12 Mb (Alonso-Blanco et al., 2005). Crossing individuals within the F2 generation for two or more subsequent generations in advanced recombinant inbred line populations (AIL; Figure 6B), increases the number of recombinations. More recombination events decreases the size of the recombination fragments and thus breaks the haplotype blocks, which increases the accuracy of the mapping (Balasubramanian et al., 2009). To increase the genetic variation in the mapping population, multiple parents have been used to create multiparent RILs, such as the Arabidopsis Multiparent RILs (AMPRIL; Huang et al., 2011a) and the Multiparent Advanced Genetic Intercross (MAGIC) lines with respectively 8 and 19 founder lines (Kover et al., 2009). These populations also go through multiple generations of recombination (MAGIC: four generations; AMPRIL: five

generations) in order to increase the mapping accuracy (Figure 6C). The larger number of segregating alleles and the higher degree of recombination led to the possibility of mapping regions of less than 1 MB.

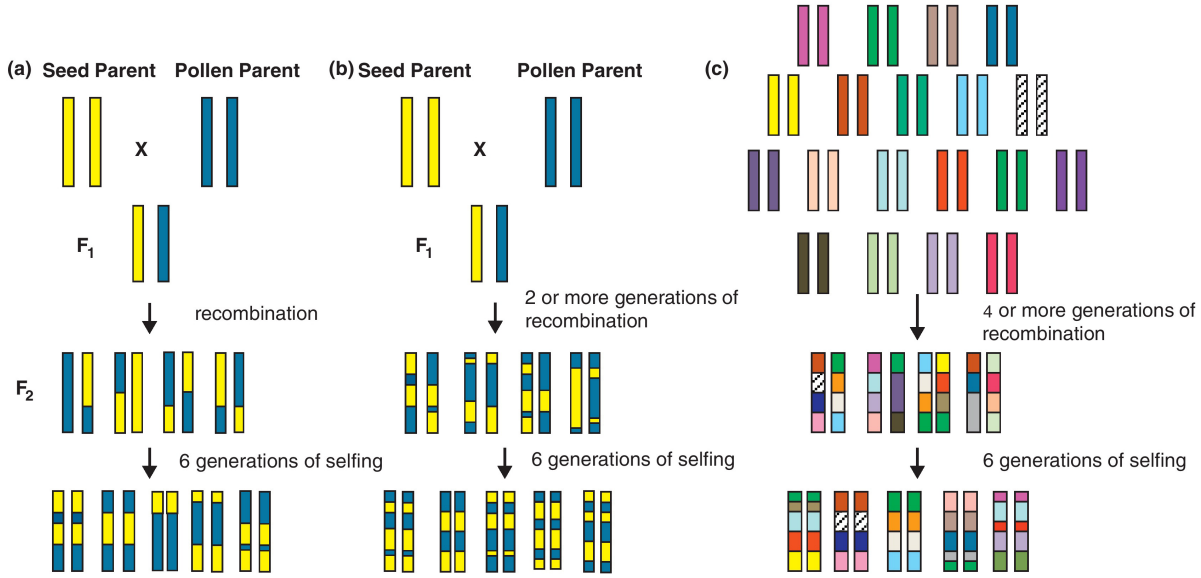


Figure 6: Creation of mapping populations. A: A typical recombinant inbred line (RIL) population, produced from the F2 of two natural accessions. B: creation of advanced intercrossed line (AIL) population, with multiple generations of recombination to reduce the size of recombination fragments. C: creation of a multiparent RIL population such as AMPRIL or MAGIC, starting from more than two parents to increase the genetic diversity and using four or more generations of recombination to reduce the size of recombination fragments. Figure derived from Kover and Mott, (2012).

Multiparent populations were already successful in mapping QTL for seed size and number (Gnan et al., 2014), flowering time (Ehrenreich et al., 2009; Salome et al., 2011), and awn presence and earliness in a wheat MAGIC population (Mackay et al., 2014; Thépot et al., 2015). Nevertheless, most QTL mapping studies are using biparental RIL populations.

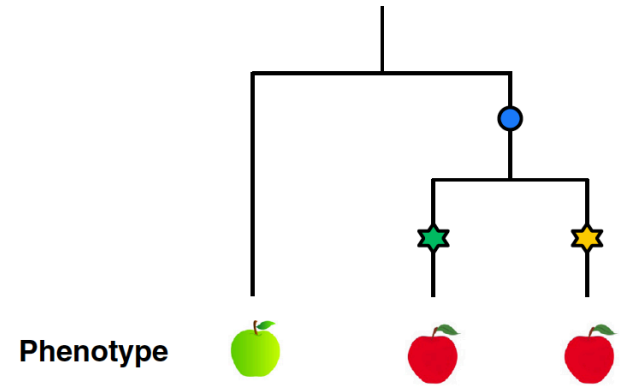
C.4 Genome-wide association mapping (GWAS)

Instead of using artificially created populations for determining the genetic architecture of specific trait, it is also possible to use the global genetic diversity for association mapping by screening collections of natural accessions in genome wide association studies (GWAS). Initially this method was developed to study traits in humans, obviously since it is impossible to create artificial populations. Arabidopsis accessions, which are highly homozygous due to selfing and which show a similar degree of linkage disequilibrium as is observed in humans, were found to be ideal for GWAS (Atwell et al., 2010). Moreover, 1300 accessions have initially been genotyped by using 250k SNP chips, which delivered the genotypic data to conduct GWAS analyses (Horton et al., 2012). Currently the 1001 genomes project is releasing genome-wide sequencing data on more than 1100 accessions, delivering more detailed genotypic data to the scientific community.

Whereas for QTL-mapping a collection will be created based on the phenotypes of the parents, for GWAS an entire collection needs to be composed. The number of accessions and the choice of accessions can influence the power and the outcome of the analysis. Below the influence of the collection composition on the GWAS outcome will be discussed.

The number of accessions will define how many repeats of each allele are present in the collection. A higher number of accessions will increase the chance to have larger groups of accessions for each allele at a given SNP (haplogroups). Larger haplogroups will result in more statistical power for the correlation between the allele and the phenotype and consequently SNPs with smaller effect sizes can associate significantly. The number of accessions required for successful GWAS will thus depend on the size of the phenotypic effect that is ought to be biologically relevant. In complex quantitative traits, where many small effect loci are involved, a sufficient amount of power is required. Simulations have shown that an increase in population size can significantly increase the power to detect small effect loci (Korte and Farlow, 2013). Hence, the required number of accessions for a successful GWAS will depend on the complexity of the trait under study. The landmark study for GWAS in Arabidopsis was successful in finding significant associations for 107 phenotypes by only using 96 accessions (Atwell et al., 2010). Also quantitative traits such as root growth parameters have successfully been studied in 96 accessions (Gifford et al., 2013; Rosas et al., 2013). However, more and more GWAS studies in Arabidopsis are using larger collections of several hundreds of accessions (Brachi et al., 2010; Li et al., 2010; Angelovici et al., 2013; Brachi et al., 2013a; Brachi et al., 2013b; Meijón et al., 2014; Slovak et al., 2014).

A



B

individual	phenotype	blue	green	orange
#1		0	0	0
#2		1	1	0
#3		1	0	1

Figure 7: Synthetic associations due to genetic heterogeneity. A: a fictive phylogenetic tree with two recent mutations (green and orange) causing a change in phenotype (red fruit) and an ancient mutation (blue) non-causal for fruit color. B: Table of the different alleles showing perfect association between fruit color and the blue allele. Figure derived from Korte and Farlow (2013).

In addition to the population size, also the choice of the accessions can make a difference in the GWAS results. In order to maximize the genetic variation, selecting geographically distinct accessions is favorable. However, this approach increases the chance to induce genetic heterogeneity. In this case different loci may underlie the trait of interest in the different accessions. If insufficient repeats of the different genetic variants of each of these loci are present in the collection, the correlation with the trait is weakened and may not be found. Moreover, a non-causative marker that is linked with both heterogeneic alleles might show a stronger association than the causative alleles (Figure 7). These kind of synthetic associations result in false positive associations. Genetic heterogeneity can be decreased by analyzing a collection with geographically less distinct accessions. However, this decreases the genetic variation in the

collection. The choice of accessions will hence be a trade-off between genetic variation and genetic heterogeneity (reviewed in Korte and Farlow, 2013).

Independent of the choice of accessions, there will always be a certain degree of relatedness between the different accessions. In other words, working with natural accessions will always lead to a certain population structure in the collections. This relatedness accounts for both causal and non-causal alleles. Above we showed that sharing non-causal alleles might lead to synthetic associations. These synthetic associations show that the non-causal alleles may have a confounding effect on the association of causal loci and the more loci that each haplogroup of the causal SNP have in common, the more false positives may arise. As more related accessions will have more loci in common, the relatedness or the population structure is a confounding factor in GWAS. Therefore it is essential to correct for population structure in the model. To this end, relatedness between the accessions is estimated from the genetic similarity (Vilhjálmsson and Nordborg, 2012). The relatedness is then, in the form of a kinship matrix, added to the mixed model. This correction for population structure significantly reduces the amount of false positive associations (Zhao et al., 2007).

Box 1: GWAS and QTL methodology.

In order to perform GWAS, two types of data need to be gathered. For the chosen mapping population the chosen trait needs to be phenotyped in each accession. And for each accession the genotype needs to be known. The first GWAS analyses in Arabidopsis made use of the 250k SNP data (Horton et al., 2012). But future studies are expected to make use of the full genome sequence data on each accession provided by the 1001genomes project (www.1001genomes.org). The genotype (G) and phenotype (y) are then associated by using a linear mixed model (1).

$$(1)\ y \sim ß_0 + ß_1 G + K + \varepsilon$$

For each SNP the model is run independently in order to estimate $ß_1$*, the effect of the genotype on the phenotype.* $ß_0$ *represents the population mean of the phenotype and the kinship matrix (K) is added as a random factor to the model in order to correct for the population structure, a known confounding factor in GWAS (Vilhjálmsson et al., 2012). When multiple correlated phenotypes are determined, for example by measuring the same trait under different environmental conditions, the linear mixed model can be extended (2) to include the environmental factor (E) and its interaction with the genotype (GE) (Korte et al., 2012).*

$$(2)\ y \sim ß_0 + ß_1 G + ß_2 E + ß_3 GE + K + \varepsilon$$

QTL mapping, another popular method to associate phenotype and genotype, uses a similar methodology (mixed models). The main difference between both techniques is the mapping population. The use of RIL populations in QTL mapping removes the population structure. Therefore no kinship matrix needs to be included in models used for QTL mapping.

Once the collection has been chosen and the trait of interest is characterized, the GWAS finally delivers a number of SNPs associated with the trait under study with certain significance (See Box 1 for methodology). Due to the genetic architecture, as discussed above, the most significant SNP is rather the most informative and not necessarily the causal SNP. Due to linkage, all genetic elements within LD with the significant SNP are potentially shaping the trait of interest. Based on literature, expression data, presence of specific functional domains and other information it is

often possible to prioritize certain genes or genetic elements within the LD region of the SNP as plausibly causal for the observed phenotype. However, GWAS remains a correlative analysis of the genetic components of a trait and causality of the correlation still needs to be proven. Therefore, validation is obligatory to be able to draw conclusions on the role of a specific locus in shaping the phenotype. Successful functional characterizations based on GWAS studies have been performed in Arabidopsis for several traits such as five branched-chain amino acid levels in seeds (Angelovici et al., 2013), ß-carotene levels in seeds (Gonzalez-Jorge et al., 2014), seed dormancy (Yano et al., 2013), leaf necrosis (Todesco et al., 2010), bacterial (Xcc568 strain) resistance (Huard-Chauveau et al., 2013), leaf sodium accumulation (Baxter et al., 2010), foliar cadmium accumulation (Chao et al., 2012) and low water potential-induced proline accumulation (Verslues et al., 2014). GWAS was also successfully performed on more complex traits such as root system architecture (Rosas et al., 2013), root cellular traits (Meijón et al., 2014), several root traits under differing nitrogen conditions (Gifford et al., 2013) and root growth related traits under standard conditions (Slovak et al., 2014). Besides Arabidopsis, GWAS studies have been carried out successfully in maize, rice, soybean, tomato and wheat (Kump et al., 2011; Huang et al., 2011b; Ranc et al., 2012; Bastien et al., 2014; Zanke et al., 2014). With the increased availability of genome-wide sequencing data, GWAS studies can be expected for more plant species in the near future.

C.5 Mapping the interaction with the environment

In addition to performing GWAS under one standard condition, it may also be interesting to analyze a trait in contrasting environments, given the importance of several environmental perturbations in natural selection. In different environments, contrasting genes may shape the same trait. To unravel this so-called genotype by environment (G×E) interaction, the phenotype of genetically identical individuals is characterized under distinct conditions. A multi-trait mixed model (MTMM) has been proposed for associating correlated traits (i.e. the same trait in different environments) with the genotype (Korte et al., 2012). This model allows for detecting loci that are specifically involved in the differential response of a trait to the different environments and hence unravel the G×E interaction.

C.6 Combining GWAS and QTL-mapping

For both QTL-mapping and GWAS, numerous studies have proven their use to find the genetics underlying a given trait. It is hard to favor one over the other method since both techniques have advantages and disadvantages and can actually profit from each other's strengths in a complementary approach. GWAS can map at a resolution that is much higher than QTL-mapping and the natural accessions cover much more genetic variation than in the average RIL population. The drawbacks of using natural accessions are the existence of a population structure, which may cause false-positive associations, and the presence of rare alleles, which may not be detected because of their low allele frequency in the population. The population structure is broken in RIL populations, resulting in fewer false-positive associations. Moreover, the rare alleles present in the parents increase in frequency in the RIL population, allowing for significant association.. However, due to the limited number of recombination events in the creation of RIL populations, the loci are mapped less accurate compared to GWAS. Therefore, QTL-mapping

could show which of the GWAS associations are true positives. In a complementary approach, the population used for the QTL-mapping will be an important choice, since both contrasting alleles for the loci detected by GWAS need to be present. The GWAS can give an indication on which contrasting accessions can be used to develop the best RIL population for the specific loci. To increase the number of contrasting alleles for different loci, in other words increase the genetic diversity, the multi-parent populations AMPRIL or MAGIC can be used (Kover and Mott, 2012). Complementing GWAS and QTL-mapping approaches were proven to be successful for flowering time in Arabidopsis (Zhao et al., 2007; Brachi et al., 2010) and are a promising combination for finding true associations between genotype and phenotype.

D. Natural variation in gene expression

In order to affect a phenotype, genetic variants can change the function of a gene when located in the coding sequence or they can have an effect on the expression of a particular gene when located in a *cis* (e.g. promoter sequence) or in a *trans* regulatory element (e.g. a transcription factor). Different micro-array and RNA-sequencing studies are shedding light on the variation in gene expression between different accessions. Whole seedling RNA-sequencing of 19 accessions showed that 46% of the expressed protein-coding genes were differentially expressed between at least two accessions (Gan et al., 2011), indicating that selection affects the transcriptional regulation. The clearest differences in expression between the accessions were noted for genes linked to biotic stress responses, whereas housekeeping genes showed much lower fold changes between accessions and often only between a limited number of accessions. So expression variation seems to be depending on the function of the genes and is likely created by the selective pressure on that specific function.

The high level of diversification in biotic stress responses is also exemplified by treating different accessions with salicylic acid (SA), one of the key hormones in the response against pathogens (Gaffney et al., 1993; Delaney et al., 1995). Among seven accessions, more than 3000 genes showed variation in their expression response upon SA. The expression variation reshaped entire gene networks in the different accessions, indicating that the variation may be present in a number of upstream transcriptional regulators (van Leeuwen et al., 2007). In response to abiotic stress such as drought, different accessions have developed different transcriptional responses. Over 4000 genes were found to be differing in their expression response to drought between accessions (Marais et al., 2012). Also for the auxin responses, involved in a plethora of plant traits, a substantial amount of gene expression variation has been observed (Delker et al., 2010). Moreover, the variation in auxin response was mainly due to expression variation whereas very few polymorphisms in coding sequences of the genes involved in these responses were observed. These findings show that different selective pressures have driven the transcriptional regulation to adapt and that gene expression differences may, in some cases, shape the phenotype to a larger extent than differences in protein function.

D.1 Mapping regulatory loci for gene expression

With the presence of substantial variation in gene expression between different accessions, the question of what is causing this variation remains. Answering this question will also indicate how

gene expression is being regulated and may provide important hints towards the key determinants of a specific phenotype, which could represent interesting candidates for further detailed investigation of the genetic architecture of a certain phenotype. Finding the causal variation can also provide information on how natural selection is shaping local adaptations and what the importance of expression variation is in certain evolutionary adaptations. By performing so-called expression QTL (eQTL) mapping it is possible to associate gene expression differences with genetic polymorphisms (reviewed by Kliebenstein, 2009 and Cubillos et al., 2014). This is done in a similar manner as QTL mapping, but gene expression levels are used as phenotype. Genome-wide expression profiles are determined by using microarrays or RNA-sequencing. The latter is preferred when using natural variation since it does not rely on probes based on a reference genome and thus alleviates the bias towards the reference genome.

Based on their location, the identified eQTLs can be classified as local eQTLs or distant eQTLs. A local eQTL can then be either a *cis* or *trans* regulatory element. The *cis* regulatory elements are genetic elements that have a physical connection with the gene they regulate and are typically located in promoter elements such as transcription factor binding domains and enhancers or silencers. Local *trans* regulatory elements can be neighboring genes with a regulatory function. The distant eQTLs, which are always *trans*, are mostly thought of as variations in transcription factors. However, *trans* regulators in yeast showed a wide array of functions and were not enriched for transcription factors (Yvert et al., 2003). Instead, the *trans* regulators encoded RNA-binding proteins, members of signaling cascades and modifiers of nucleosome composition (reviewed by Albert and Kruglyak, 2015) so *trans* regulatory elements can be part of a wide array of regulatory strategies. Many of the *trans* eQTLs are clustered in hotspots, most likely caused by one master regulator that regulates many genes, although it cannot be excluded that different regulatory loci are located together (Cubillos et al., 2012).

Different studies have found a larger extent of *cis* regulators in comparison to the detected *trans* regulators (Keurentjes et al., 2007; West et al., 2007; Zhang et al., 2011). Loci found in *trans* are however often showing a smaller phenotypic effect compared to the *cis* located loci and thus require more statistical power to be detected. Therefore, studies with differing statistical power may detect different amounts of *trans* loci. Moreover, mapping populations may differ in the amount of *cis* genetic variation (reviewed by Kliebenstein, 2009). Despite their small phenotypic effect, *trans* regulators play a significant role in response to distinct environments or in specific cell types, whereas *cis* regulators show a more robust response over different conditions (Dimas et al., 2009; Drost et al., 2010; Barriere et al., 2011; Grundberg et al., 2011). The *cis* located regulators are probably involved in more general gene expression regulatory mechanisms, while the *trans* regulators are involved in the regulation of transcription in a specific environment, cell type or developmental phase.

E. Future perspectives for association mapping

With the emergence of RNA-sequencing, it becomes feasible to characterize the transcriptome of different natural variants of many different species. Together with projects such as the sequencing of 1001 Arabidopsis accessions and the methods discussed above, we expect a growing number of association studies on gene expression. These analyses will increase our

understanding of gene expression regulation and has the potential to determine specific regulatory interactions in different tissues and environments. Knowledge on the different loci influencing gene expression will ultimately lead to creating new or fine-tuning existing gene regulatory networks. The integration of the regulatory interactions with phenotypic associations from GWAS studies may aid to distinguish false from true positive associations. Integration of phenotype and expression associations will shed light on the transcriptional regulation of certain traits. Moreover, the possibilities of association mapping do not stop at gene expression. Similar approaches may be conducted on the different regulatory levels. Association mapping of epigenome, proteome and metabolome data can further fine-tune the knowledge on how genetics are influencing the phenotype. Ultimately the integration of these different data types may give clear sight on genotype to phenotype relations and their interaction with the environment.

References

Achard P, Gusti A, Cheminant S, Alioua M, Dhondt S, Coppens F, Beemster GTS, Genschik P (2009) Gibberellin signaling controls cell proliferation rate in Arabidopsis. Curr Biol 19: 1188–1193

Aguirrezabal L, Bouchier-Combaud S, Radziejwoski A, Dauzat M, Cookson SJ, Granier C (2006) Plasticity to soil water deficit in Arabidopsis thaliana: dissection of leaf development into underlying growth dynamic and cellular variables reveals invisible phenotypes. Plant Cell Environ 29: 2216–2227

Albert FW, Kruglyak L (2015) The role of regulatory variation in complex traits and disease. Nature Review Genetics 16: 197–212

Alonso-Blanco C, Koornneef M (2000) Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics. Trends Plant Sci 5: 22–29

Alonso-Blanco C, Méndez-Vigo B, Koornneef M (2005) From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. Int J Dev Biol 49: 717–732

Alves MS, Fontes E, Fietto LG (2011) EARLY RESPONSIVE to DEHYDRATION 15, a new transcription factor that integrates stress signaling pathways. Plant Signal Behav. doi: 10.4161/psb.6.12.18268

Angelovici R, Lipka AE, Deason N, Gonzalez-Jorge S, Lin H, Cepela J, Buell R, Gore MA, Dellapenna D (2013) Genome-wide analysis of branched-chain amino acid levels in Arabidopsis seeds. Plant Cell 25: 4827–4843

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815

Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, et al (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627–631

Baerenfaller K, Massonnet C, Walsh S, Baginsky S, Bühlmann P, Hennig L, Hirsch-Hoffmann M, Howell KA, Kahlau S, Radziejwoski A, et al (2012) Systems-based analysis of Arabidopsis leaf growth reveals adaptation to water deficit. Mol Syst Biol 8: 606

Bahrun A, Jensen CR, Asch F, Mogensen VO (2002) Drought-induced changes in xylem pH, ionic composition, and ABA concentration act as early signals in field-grown maize (*Zea mays* L.). J Exp Bot 53: 251–263

Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, Maloof JN, Loudet O, Trainer GT, Dabi T, Borevitz JO, et al (2009) QTL mapping in new *Arabidopsis thaliana* advanced intercross-recombinant inbred lines. PLoS ONE. doi: 10.1371/journal.pone.0004318

Barriere A, Gordon KL, Ruvinsky I (2011) Distinct Functional Constraints Partition Sequence Conservation in a cis-Regulatory Element. PLoS Genet. doi: 10.1371/journal.pgen.1002095

Bastien M, Sonah H, Belzile F (2014) Genome Wide Association Mapping of Sclerotinia sclerotiorum Resistance in Soybean with a Genotyping-by-Sequencing Approach. The Plant Genome 7: 1–13

Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E, Li Y, Bergelson J, Borevitz JO, Nordborg M, et al

(2010) A Coastal Cline in Sodium Accumulation in Arabidopsis thaliana Is Driven by Natural Variation of the Sodium Transporter AtHKT1;1. PLoS Genet 6: e1001193

Bergmann DC, Sack FD (2007) Stomatal development. Annu Rev Plant Biol 58: 163–181

Brachi B, Faure N, Bergelson J, Cuguen J, Roux F (2013a) Genome-wide association mapping of flowering time in Arabidopsis thalianain nature: genetics for underlying components and reaction norms across two successive years. Acta Botanica Gallica 160: 205–219

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F (2010) Linkage and association mapping of Arabidopsis thaliana flowering time in nature. PLoS Genet 6: e1000940

Brachi B, Villoutreix R, Faure N, Hautekèete N, Piquot Y, Pauwels M, Roby D, Cuguen J, Bergelson J, Roux F (2013b) Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in Arabidopsis thaliana. Molecular Ecology 22: 4222–4240

Burssens S, Himanen K, van de Cotte B, Beeckman T, Van Montagu M, Inz x000E9 D, Verbruggen N (2000) Expression of cell cycle regulatory genes and morphological alterations in response to salt stress in *Arabidopsis thaliana*. Planta 211: 632–640

Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, et al (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat Genet 43: 956–963

Chao D-Y, Silva A, Baxter I, Huang YS, Nordborg M, Danku J, Lahner B, Yakubova E, Salt DE (2012) Genome-wide association studies identify heavy metal ATPase3 as the Primary Determinant of Natural Variation in Leaf Cadmium in *Arabidopsis thaliana*. PLoS Genet. doi: 10.1371/journal.pgen.1002923

Claeys H, De Bodt S, Inzé D (2014a) Gibberellins and DELLAs: central nodes in growth regulatory networks. Trends Plant Sci 19: 231–239

Claeys H, Inzé D (2013) The agony of choice: how plants balance growth and survival under water-limiting conditions. Plant Physiol 162: 1768–1779

Claeys H, Skirycz A, Maleux K, Inzé D (2012) DELLA signaling mediates stress-induced cell differentiation in Arabidopsis leaves through modulation of anaphase-promoting complex/cyclosome activity. Plant Physiol 159: 739–747

Claeys H, Van Landeghem S, Dubois M, Maleux K, Inzé D (2014b) What is Stress? Dose-Response Effects in Commonly Used In Vitro Stress Assays. Plant Physiol 165: 519–527

Clauw P, Coppens F, De Beuf K, Dhondt S, Van Daele T, Maleux K, Storme V, Clement L, Gonzalez N, Inzé D (2015) Leaf responses to mild drought stress in natural variants of Arabidopsis. Plant Physiol 167: 800–816

Cubillos FA, Coustham V, Loudet O (2012) Lessons from eQTL mapping studies: non-coding regions and their role behind natural phenotypic variation in plants. Curr Opin Plant Biol 15: 192–198

Cubillos FA, Stegle O, Grondin C, Canut M, Tisné S, Gy I, Loudet O (2014) Extensive cis-regulatory variation robust to environmental perturbation in Arabidopsis. Plant Cell 26: 4298–4310

Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR (2010) Abscisic Acid: Emergence of a Core Signaling Network. Annu Rev Plant Biol 61: 651–679

Dat J, Vandenabeele S, Vranová E, Van Montagu M, Inze D, Van Breusegem F (2000) Dual action of the active oxygen species during plant stress responses. Cell Mol Life Sci 57: 779–795

De Veylder L, Beeckman T, Inzé D (2007) The ins and outs of the plant cell cycle. Nat Rev Mol Cell Biol 8: 655–665

Delaney TP, Friedrich L, Ryals JA (1995) Arabidopsis signal transduction mutant defective in chemically and biologically induced disease resistance. Proc Natl Acad Sci USA 92: 6602–6606

Delker C, Pöschl Y, Raschke A, Ullrich K, Ettingshausen S, Hauptmann V, Grosse I, Quint M (2010) Natural variation of transcriptional auxin response networks in *Arabidopsis thaliana*. Plant Cell 22: 2184–2200

Dhondt S, Wuyts N, Inzé D (2013) Cell to whole-plant phenotyping: the best is yet to come. Trends Plant Sci 18: 428–439

Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Arcelus MG, Sekowska M, et al (2009) Common Regulatory Variation Impacts Gene Expression in a Cell Type-Dependent Manner. Science 325: 1246–1250

Döll P, Siebert S (2002) Global modeling of irrigation water requirements. Water Resour Res. doi: 10.1029/2001WR000355

Drost DR, Benedict CI, Berg A, Novaes E, Novaes CRDB, Yu Q, Dervinis C, Maia JM, Yap J, Miles B, et al (2010) Diversification in the genetic architecture of gene expression and transcriptional networks in organ differentiation of Populus. Proc Natl Acad Sci USA 107: 8492–8497

Dubois M, Skirycz A, Claeys H, Maleux K, Dhondt S, De Bodt S, Vanden Bossche R, De Milde L, Yoshizumi T, Matsui M, et al (2013) ETHYLENE RESPONSE FACTOR6 acts as a central regulator of leaf growth under water-limiting conditions in Arabidopsis. Plant Physiol 162: 319–332

Ehrenreich IM, Hanzawa Y, Chou L, Roe JL, Kover PX, Purugganan MD (2009) Candidate gene association mapping of Arabidopsis flowering time. Genetics 183: 325–335

FAO (2011) FAO's preparedness and response to the Horn of Africa drought. Retrieved from http://www.fao.org/fileadmin/user_upload/emergencies/docs/FAO_ongoing_response_to_the_drought_in_the_Horn_of_Africa.pdf

Finkelstein R (2013) Abscisic Acid synthesis and response. The Arabidopsis Book 11: e0166

Fischer RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the royals society of Edinburgh 52: 399–433

Fricke W (2006) The short-term growth response to salt of the developing barley leaf. J Exp Bot 57: 1079–1095

Fujii H, Chinnusamy V, Rodrigues A, Rubio S, Antoni R, Park S-Y, Cutler SR, Sheen J, Rodriguez PL, Zhu J-K (2009) In vitro reconstitution of an abscisic acid signalling pathway. Nature 462: 660–664

Fujita Y, Fujita M, Shinozaki K, Yamaguchi-Shinozaki K (2011) ABA-mediated transcriptional regulation in response to osmotic stress in plants. J Plant Res 124: 509–525

Gaffney T, Friedrich L, Vernooij B, Negrotto D, Nye G, Uknes S, Ward E, Kessmann H, Ryals J (1993) Requirement of salicylic Acid for the induction of systemic acquired resistance. Science 261: 754–756

Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT, et al (2011) Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. 477: 419–423

Gechev TS, Dinakar C, Benina M, Toneva V, Bartels D (2012) Molecular mechanisms of desiccation tolerance in resurrection plants. Cell Mol Life Sci 69: 3175–3186

Geisler M, Nadeau J, Sack FD (2000) Oriented asymmetric divisions that generate the stomatal spacing pattern in arabidopsis are disrupted by the too many mouths mutation. Plant Cell 12: 2075–2086

Gifford ML, Banta JA, Katari MS, Hulsmans J, Chen L, Ristova D, Tranchina D, Purugganan MD, Coruzzi GM, Birnbaum KD (2013) Plasticity Regulators Modulate Specific Root Traits in Discrete Nitrogen Environments. PLoS Genet. doi: 10.1371/journal.pgen.1003760

Gnan S, Priest A, Kover PX (2014) The Genetic Basis of Natural Variation in Seed Size and Seed Number and Their Trade-Off Using Arabidopsis thaliana MAGIC Lines. Genetics. doi: 10.1534/genetics.114.170746/-/DC1

Gonzalez N, Vanhaeren H, Inzé D (2012) Leaf size control: complex coordination of cell division and expansion. Trends Plant Sci 17: 332–340

Gonzalez-Jorge S, Ha SH, Magallanes-Lundback M, Gilliland LU, Zhou A, Lipka AE, Nguyen YN, Angelovici R, Lin H, Cepela J, et al (2014) *CAROTENOID CLEAVAGE DIOXYGENASE4* is a negative regulator of ß-carotene content in *Arabidopsis* Seeds. Plant Cell 25: 4812–4826

Granier C, Inze D, Tardieu F (2000) Spatial distribution of cell division rate can be deduced from that of p34(cdc2) kinase activity in maize leaves grown at contrasting temperatures and soil water conditions. Plant Physiol 124: 1393–1402

Griffiths CA, Gaff DF, Neale AD (2014) Drying without senescence in resurrection plants. Front Plant Sci. doi: 10.3389/fpls.2014.00036

Grundberg E, Adoue V, Kwan T, Ge B, Duan QL, Lam KCL, Koka V, Kindmark A, Weiss ST, Tantisira K, et al (2011) Global Analysis of the Impact of Environmental Perturbation on cis-Regulation of Gene Expression. PLoS Genet. doi: 10.1371/journal.pgen.1001279

Harb A, Krishnan A, Ambavaram M, Pereira A (2010) Molecular and physiological analysis of drought stress in Arabidopsis reveals early responses leading to acclimation in plant growth. Plant Physiol 154: 1254–1271

Hoffmann MH (2002) Biogeography of Arabidopsis thaliana (L.) Heynh. (Brassicaceae). J Biogeogr 29: 125–134

Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Muliyati NW, Platt A, Sperone FG, Vilhjálmsson BJ, et al (2012) Genome-wide patterns of genetic variation in worldwide. Nat Genet 44: 212–216

Huang X, Paulo M-J, Boer M, Effgen S, Keizer P, Koornneef M, van Eeuwijk FA (2011a) Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. Proc Natl Acad Sci USA 108: 4488–4493

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, et al (2011b) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nature Publishing Group 44: 32–39

Huard-Chauveau C, Perchepied L, Debieu M, Rivas S, Kroj T, Kars I, Bergelson J, Roux F, Roby D (2013) An Atypical Kinase under Balancing Selection Confers Broad-Spectrum Disease Resistance in Arabidopsis. PLoS Genet. doi: 10.1371/journal.pgen.1003766

Hundertmark M, Hincha DK (2008) LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. BMC Genomics 9: 118–122

Ingram J, Bartels D (1996) The molecular basis of dehydration tolerance in plants. Annu Rev Plant Physiol 47: 377–403

IPCC (2014) Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1132 pp.

Kariola T, Brader G, Helenius E, Li J, Heino P, Palva ET (2006) EARLY RESPONSIVE TO DEHYDRATION 15, a negative regulator of abscisic acid responses in Arabidopsis. Plant Physiol 142: 1559–1573

Keurentjes JJB, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJM, Vreugdenhil D, Koornneef M, Jansen RC (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proc Natl Acad Sci USA 104: 1708–1713

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and linkage disequilibrium in Arabidopsis thaliana. Nat Genet 39: 1151–1155

Kiyosue T, Abe H, Yamaguchi-Shinozaki K, Shinozaki K (1998) ERD6, a cDNA clone for an early dehydration-induced gene of Arabidopsis, encodes a putative sugar transporter1The nucleotide sequence reported in this paper has been submitted to DDBJ with the accession number of D89051.1. *Biochimica et Biophysica Acta* (BBA) - Biomembranes 1370: 187–191

Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K (1994a) Characterization of two cDNAs (ERD10 and ERD14) corresponding to genes that respond rapidly to dehydration stress in Arabidopsis thaliana. Plant Cell Physiol 35: 225–231

Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K (1993) Characterization of two cDNAs (ERD11 and ERD13) for dehydration-inducible genes that encode putative glutathione S-transferases in *Arabidopsis thaliana* L. FEBS Letters 335: 189–192

Kiyosue T, YAMAGUCHISHINOZAKI K, Shinozaki K (1994b) Cloning of cdnas for genes that are early-responsive to dehydration stress (ERDs) in *Arabidopsis thaliana* L - Identification of 3 ERDs as Hsp cognate genes. Plant Mol Biol 25: 791–798

Kiyosue T, Yoshiba Y, Yamaguchi-Shinozaki K, Shinozaki K (1996) A nuclear gene encoding mitochondrial proline dehydrogenase, an enzyme involved in proline metabolism, is upregulated by proline but downregulated by dehydration in Arabidopsis. Plant Cell 8: 1323–1335

Kliebenstein D (2009) Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. Annu Rev Plant Biol 60: 93–114

Kline KG, Barrett-Wilt GA, Sussman MR (2010) In planta changes in protein phosphorylation induced by the plant hormone abscisic acid. PNAS 107: 15986–15991

Komaki S, Sugimoto K (2012) Control of the plant cell cycle by developmental and environmental cues. Plant Cell Physiol 53: 953–964

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in *Arabidopsis thaliana*. Annu Rev Plant Biol 55: 141–172

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9: 29

Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M (2012) A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat Genet 44: 1066–1071

Kosma DK, Bourdenx B, Bernard A, Parsons EP, Lü S, Joubès J, Jenks MA (2009) The impact of water deficiency on leaf cuticle lipids of Arabidopsis. Plant Physiol 151: 1918–1929

Kover PX, Mott R (2012) Mapping the genetic basis of ecologically and evolutionarily relevant traits in *Arabidopsis thaliana*. Curr Opin Plant Biol 15: 212–217

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R (2009) A Multiparent Advanced Generation Inter-Cross to Fine-Map Quantitative Traits in Arabidopsis thaliana. PLoS Genet 5: e1000551–15

Krämer PJ, Boyer JS (1995) Water relations of plant and soils. Water relations of plant and soils

Krämer U (2015) Planting molecular functions in an ecological context with *Arabidopsis thaliana*. eLife. doi: 10.7554/eLife.06100.001

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, et al (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43: 163–168

Leung J, Bouvier-Durand M, Morris PC, Guerrier D, Chefdor F, Giraudat J (1994) Arabidopsis ABA response gene ABI1: features of a calcium-modulated protein phosphatase. Science 264: 1448–1452

Leung J, Merlot S, Giraudat J (1997) The Arabidopsis ABSCISIC ACID-INSENSITIVE2 (ABI2) and ABI1 genes encode homologous protein phosphatases 2C involved in abscisic acid signal transduction. Plant Cell 9: 759–771

Li Y, Huang Y, Bergelson J, Nordborg M, Borevitz JO (2010) Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. PNAS 107: 21199–21204

Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1998) Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis. Plant Cell 10: 1391–1406

Mackay IJ, Bansept-Basler P, Barber T (2014) An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. G3: Genes| Genomes| …. doi: 10.1534/g3.114.012963/-/DC1

Marais Des DL, McKay JK, Richards JH, Sen S, Wayne T, Juenger TE (2012) Physiological genomics of response to soil drying in diverse *Arabidopsis* accessions. Plant Cell 24: 893–914

Meijón M, Satbhai SB, Tsuchimatsu T, Busch W (2014) Genome-wide association study using cellular traits identifies a new regulator of root development in Arabidopsis. Nat Genet 46: 77–81

Meyer K, Leube MP, Grill E (1994) A protein phosphatase 2C involved in ABA signal transduction in Arabidopsis thaliana. Science 264: 1452–1455

Michelena VA, Boyer JS (1982) Complete turgor maintenance at low water potentials in the elongating region of maize leaves. Plant Physiol 69: 1145–1149

Mitra SK, Gantt JA, Ruby JF, Clouse SD, Goshe MB (2007) Membrane proteomic analysis of Arabidopsis thaliana using alternative solubilization techniques. J Proteome Res 6: 1933–1950

Mittler R, Vanderauwera S, Suzuki N, Miller G, Tognetti VB, Vandepoele K, Gollery M, Shulaev V, Van Breusegem F (2011) ROS signaling: the new wave? Trends Plant Sci 16: 300–309

Miyakawa T, Fujita Y, Yamaguchi-Shinozaki K, Tanokura M (2013) Structure and function of abscisic acid receptors. Trends Plant Sci 18: 259–266

Moore JP, Vicré-Gibouin M, Farrant JM, Driouich A (2008) Adaptations of higher plant cell walls to water loss: drought vs desiccation. Physiol Plant 134: 237–245

Muller B, Pantin F, Génard M, Turc O, Freixes S, Piques M, Gibon Y (2011) Water deficits uncouple growth from photosynthesis, increase C content, and modify the relationships between C and growth in sink organs. J Exp Bot 62: 1715–1729

Murphy R, Ortega J (1995) A New Pressure Probe Method to Determine the Average Volumetric Elastic Modulus of Cells in Plant Tissue. Plant Physiol 107: 995–1005

Nakashima K, Satoh R, Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K (1998) A gene encoding proline dehydrogenase is not only induced by proline and hypoosmolarity, but is also developmentally regulated in the reproductive organs of Arabidopsis. Plant Physiol 118: 1233–1241

Nakashima K, Shinwari ZK, Sakuma Y, Seki M, Miura S, Shinozaki K, Yamaguchi-Shinozaki K (2000) Organization and expression of two Arabidopsis DREB2 genes encoding DRE-binding proteins involved in dehydration- and high-salinity-responsive gene expression. Plant Mol Biol 42: 657–665

Narusaka Y, Nakashima K, Shinwari ZK, Sakuma Y, Furihata T, Abe H, Narusaka M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of Arabidopsis rd29A gene in response to dehydration and high-salinity stresses. Plant J 34: 137–148

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, et al (2005) The Pattern of Polymorphism in Arabidopsis thaliana. Plos Biol 3: e196

Ohkuma K, Lyon JL, Addicott FT, Smith OE (1963) Abscisin II, an abscission-accelerating substance from young cotton fruit. Science 142: 1592–1593

Parent B, Hachez C, Redondo E, Simonneau T, Chaumont F, Tardieu F (2009) Drought and abscisic acid effects on aquaporin content translate into changes in hydraulic conductivity and leaf growth rate: a trans-scale approach. Plant Physiol 149: 2000–2012

Park S-Y, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow T-FF, et al (2009) Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. Science 324: 1068–1071

Parry MA, Andralojc PJ, Scales JC, Salvucci ME, Carmo-Silva AE, Alonso H, Whitney SM (2013) Rubisco activity and regulation as targets for crop improvement. J. Exp. Bot. 64: 717–730

Peres A, Churchman ML, Hariharan S, Himanen K, Verkest A, Vandepoele K, Magyar Z, Hatzfeld Y, Van Der Schueren E, Beemster GTS, et al (2007) Novel plant-specific cyclin-dependent kinase inhibitors induced by biotic and abiotic stresses. J Biol Chem 282: 25588–25596

Pereyra-Irujo GA, Velazquez L, Lechner L, Aguirrezabal L (2008) Genetic variability for leaf growth rate and duration under water deficit in sunflower: analysis of responses at cell, organ, and plant level. J Exp Bot 59: 2221–2232

Peterson KM, Rychel AL, Torii KU (2010) Out of the mouths of plants: the molecular basis of the evolution and diversity of stomatal development. Plant Cell 22: 296–306

Pettkó-Szandtner A, Mészáros T, Horváth GV, Bakó L, Csordás-Tóth E, Blastyák A, Zhiponova M, Miskolczi P, Dudits D (2006) Activation of an alfalfa cyclin-dependent kinase inhibitor by calmodulin-like domain protein kinase. Plant J 46: 111–123

Pillitteri LJ, Dong J (2013) Stomatal development in Arabidopsis. The Arabidopsis Book 11: e0162

Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Ågren J, Bossdorf O, Byers D, Donohue K, et al (2010) The scale of population structure in *Arabidopsis thaliana*. PLoS Genet. doi: 10.1371/journal.pgen.1000843

Ranc N, Muños S, Xu J, Le Paslier M-C, Chauveau A, Bounon R, Rolland S, Bouchet J-P, Brunel D, Causse M (2012) Genome-wide association mapping in tomato (Solanum lycopersicum) is possible using genome admixture of *Solanum lycopersicum* var. cerasiforme. G3 (Bethesda) 2: 853–864

Rodrigues A, Adamo M, Crozet P, Margalha L, Confraria A, Martinho C, Elias A, Rabissi A, Lumbreras V, Gonzalez-Guzman M, et al (2013) ABI1 and PP2CA phosphatases are negative regulators of Snf1-related protein kinase1 signaling in Arabidopsis. Plant Cell 25: 3871–3884

Rodriguez PL, Benning G, Grill E (1998) ABI2, a second protein phosphatase 2C involved in abscisic acid signal transduction in Arabidopsis. FEBS Letters 421: 185–190

Rosas U, Cibrian-Jaramillo A, Ristova D, Banta JA, Gifford ML, Fan AH, Zhou RW, Kim GJ, Krouk G, Birnbaum KD, et al (2013) Integration of responses within and across Arabidopsis natural accessions uncovers loci controlling root systems architecture. Proc Natl Acad Sci USA 110: 15133–15138

Rubio S, Rodrigues A, Saez A, Dizon MB, Galle A, Kim T-H, Santiago J, Flexas J, Schroeder JI, Rodriguez PL (2009) Triple loss of function of protein phosphatases type 2C leads to partial constitutive response to endogenous abscisic acid. Plant Physiol 150: 1345–1355

Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K (2002) DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem Biophys Res Commun 290: 998–1009

Salome PA, Bomblies K, Laitinen RAE, Yant L, Mott R, Weigel D (2011) Genetic Architecture of Flowering-Time Variation in Arabidopsis thaliana. Genetics 188: 421–433

Schuppler U, He P, John P, Munns R (1998) Effect of water stress on cell division and cell-division-cycle 2-like cell-cycle kinase activity in wheat leaves. Plant Physiol 117: 667–678

Shatil-Cohen A, Attia Z, Moshelion M (2011) Bundle-sheath cell regulation of xylem-mesophyll water transport via aquaporins under drought stress: a target of xylem-borne ABA? Plant J 67: 72–80

Sheen J (1998) Mutational analysis of protein phosphatase 2C involved in abscisic acid signal transduction in higher plants. Proc Natl Acad Sci USA 95: 975–980

Skirycz A, Claeys H, De Bodt S, Oikawa A, Shinoda S, Andriankaja M, Maleux K, Eloy NB, Coppens F, Yoo S-D, et al (2011a) Pause-and-stop: the effects of osmotic stress on cell proliferation during early leaf development in Arabidopsis and a role for ethylene signaling in cell cycle arrest. Plant Cell 23: 1876–1888

Skirycz A, De Bodt S, Obata T, De Clercq I, Claeys H, De Rycke R, Andriankaja M, Van Aken O, Van Breusegem F, Fernie AR, et al (2010) Developmental stage specificity and the role of mitochondrial metabolism in the response of Arabidopsis leaves to prolonged mild osmotic stress. Plant Physiol 152: 226–244

Skirycz A, Inzé D (2010) More from less: plant growth under limited water. Curr Opin Biotechnol 21: 197–203

Skirycz A, Vandenbroucke K, Clauw P, Maleux K, De Meyer B, Dhondt S, Pucci A, Gonzalez N, Hoeberichts F, Tognetti VB, et al (2011b) Survival and growth of Arabidopsis plants given limited water are not equal. Nat Biotechnol 29: 212–214

Slovak R, Goeschl C, Su X, Shimotani K, Shiina T, Busch W (2014) A Scalable open-source pipeline for large-scale root phenotyping of *Arabidopsis*. Plant Cell 26: 2390–2403

Soitamo AJ, Piippo M, Allahverdiyeva Y, Battchikova N, Aro E-M (2008) Light has a specific role in modulating Arabidopsis gene expression at low temperature. BMC Plant Biology. doi: 10.1186/1471-2229-8-13

Stitt M, Schulze D (1994) Does Rubisco control the rate of photosynthesis and plant growth? An exercise in molecular ecophysiology, Plant Cell Environ. 17: 465–487.

Stockinger EJ, Gilmour SJ, Thomashow MF (1997) Arabidopsis thaliana CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc Natl Acad Sci USA 94: 1035–1040

Szabados L, Kovacs H, Zilberstein A, Bouchereau A (2011) Plants in Extreme Environments: Importance of Protective Compounds in Stress Tolerance. Plant Responses to Drought and Salinity Stress: Developments in a Post-Genomic Era 57: 105–150

Tardieu F, Parent B, Simonneau T (2010) Control of leaf growth by abscisic acid: hydraulic or non-hydraulic processes? Plant Cell Environ 33: 636–647

Thépot S, Restoux G, Goldringer I, Hospital F, Gouache D, Mackay I, Enjalbert J (2015) Efficiently tracking selection in a multiparental population: the case of earliness in wheat. Genetics 199: 609–623

Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M, Epple P, Kuhns C, Sureshkumar S, Schwartz C, Lanz C, et al (2010) Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. Nature 465: 632–636

Umezawa T, Nakashima K, Miyakawa T, Kuromori T, Tanokura M, Shinozaki K, Yamaguchi-Shinozaki K (2010) Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. Plant Cell Physiol 51: 1821–1839

UNESCO - United Nations Educational Scientific and Cultural Organization (2001) Securing the food supply. Paris: UNESCO.

van Leeuwen H, Kliebenstein DJ, West MAL, Kim K, van Poecke R, Katagiri F, Michelmore RW, Doerge RW, St Clair DA (2007) Natural variation among *Arabidopsis thaliana* accessions for transcriptome response to exogenous salicylic acid. Plant Cell 19: 2099–2110

Verslues PE, Agarwal M, Katiyar-Agarwal S, Zhu J, Zhu J-K (2006) Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. Plant J 45: 523–539

Verslues PE, Lasky JR, Juenger TE, Liu TW, Kumar MN (2014) Genome-wide association mapping combined with reverse genetics identifies new effectors of low water potential-induced proline accumulation in *Arabidopsis*. Plant Physiol 164: 144–159

Veselov DS, Mustafina AR, Sabirjanova IB, Akhiyarova GR, Dedov AV, Veselov SU, Kudoyarova GR (2002) Effect of PEG-treatment on the leaf growth response and auxin content in shoots of wheat seedlings. Plant Growth Regulation 38: 191–194

Veselov DS, Sharipova GV, Veselov SU, Kudoyarova GR (2008) The effects of NaCl treatment on water relations, growth, and ABA content in barley cultivars differing in drought tolerance. J Plant Growth Regul 27: 380–386

Vilhjálmsson BJ, Nordborg M (2012) The nature of confounding in genome-wide association studies. Nature Nat Rev Genet 14: 1–2

Weigel D (2012) Natural Variation in Arabidopsis: From Molecular Genetics to Ecological Genomics. Plant Physiol 158: 2–22

West MAL, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, St Clair DA (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. Genetics 175: 1441–1450

Whitney SM, Sharwood RE, Orr D, White SJ, Alonso H, Galmés J (2011) Isoleucine 309 acts as a C4 catalytic switch that increases ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) carboxylation rate in Flaveria, PNAS 108: 14688–14693

Wilkinson S, Davies WJ (2010) Drought, ozone, ABA and ethylene: new insights from cell to plant to community. Plant Cell Environ 33: 510–525

The World Bank (2012) Severe drought drive food prices higher, threatening the poor. Retrieved from http://www.worldbank.org/en/news/press-release/2012/08/30/severe-droughts-drive-food-prices-higher-threatening-poor

Yamaguchi-Shinozaki K, Shinozaki K (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu Rev Plant Biol 57: 781–803

Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN (1982) Living with water stress: evolution of osmolyte systems. Science 217: 1214–1222

Yano R, Takebayashi Y, Nambara E, Kamiya Y, Seo M (2013) Combining association mapping and transcriptomics identify HD2B histone deacetylase as a genetic factor associated with seed dormancy in Arabidopsis thaliana. Plant J 74: 815–828

Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L (2003) Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat Genet 35: 57–64

Zanke CD, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Hinze M, Neumann K, et al (2014) Whole genome association mapping of plant height in winter wheat (*Triticum aestivum* L.). PLoS ONE 9: e113287–16

Zhang J, Nguyen HT, Blum A (1999) Genetic analysis of osmotic adjustment in crop plants. J Exp Bot 50: 291–302

Zhang X, Cal AJ, Borevitz JO (2011) Genetic architecture of regulatory variation in *Arabidopsis thaliana*. Genome Research 21: 725–733

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, et al (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 3: e4

1. Survival and growth of Arabidopsis plants given limited water are not equal*

Aleksandra Skirycz[1,2,#], Korneel Vandenbroucke[1,2,#], Pieter Clauw[1,2], Katrien Maleux[1,2], Bjorn De Meyer[1,2], Stijn Dhondt[1,2], Anna Pucci[1,2,4], Nathalie Gonzalez[1,2], Frank Hoeberichts[1,2], Vanesa B Tognetti[1,2], Massimo Galbiati[3], Chiara Tonelli[3], Frank Van Breusegem[1,2], Marnik Vuylsteke[1,2] & Dirk Inzé[1,2]

[1]Department of Plant Systems Biology, VIB, Ghent, Belgium.
[2]Department of Plant Biotechnology and Genetics, Ghent University, Ghent, Belgium.
[3]Dipartimento di Scienze Biomolecolari e Biotecnologie, Università degli Studi di Milano, Milano, Italy.

AUTHOR CONTRIBUTIONS

Experimental work was performed by A.S., K.V., P.C., K.M., A.P., N.G., F.H. and V.B.T. Unpublished transgenic lines were provided by M.G. and C.T.; B.D.M. and S.D. designed and programmed WIWAM; M.V. performed the statistical analysis; and F.V.B. and D.I. supervised the project.

---

* This work was published in Nature Biotechnology 2011, vol. 29: 212-214.

# These authors contributed equally to this work

Although drought tolerance is a central concern of plant research, the translatability for crop improvement is relatively low. Here we report on a major contributing factor to this lack of success. Drought tolerance is predominately scored based on an improved survival rate under lethal conditions that, as demonstrated by our study, does not predict superior growth performance and, thus, biomass yield gain, under moderate drought often encountered in the field.

Drought tolerance is a major subject of trait research for agro biotech companies and thousands of academic papers have been published on the topic. Consequently, there is a plethora of reports on improved drought tolerance, mainly in the model plant Arabidopsis thaliana (Umezawa et al., 2006). Classic genetic engineering approaches involve target genes that function in mechanisms used by plants to avoid and/or tolerate drought, such as stomatal conductance or osmolyte production (Hirayama and Shinozaki, 2010). Such genes, frequently identified through expression profiling, include signaling components and downstream effector genes. However, despite the apparent success of stress research on model plants, rarely are the findings applied to improve crops. Only a few genes have been characterized that enhance stress tolerance in model plants or crops leading to increased yields (De Block et al., 2005; Nelson et al., 2007; Castiglioni et al., 2008; Li et al., 2008) and the molecular mechanisms through which they work remain only partly understood. One of the key reasons relates to the genetic and physiological differences between model and crop species.

In Arabidopsis research, drought tolerance is assessed predominantly under quite severe conditions in which plant survival is scored after a prolonged period of soil drying. However, in temperate climates, limited water availability rarely causes plant death, but restricts biomass and seed yield. To study the relation between survival and biomass gain under drought, we analyzed the growth of transgenic Arabidopsis plants with increased tolerance to lethal stress in a mild stress assay. An extensive literature screen was conducted to identify Arabidopsis genes that, in gain- or loss-of- function situations, confer stress tolerance in Arabidopsis, without growth penalty under control conditions. Although drought and osmotic stresses were prioritized, other related ones, including salt, heat and oxidative stresses, were considered as well. The final selection consisted of 25 genes, which we designate 'stress tolerance genes' (STGs), involved in diverse aspects of stress tolerance and in a wild-type Columbia-0 (Col-0) background (Table 1). We added two additional lines (MYB90 and tAPX) that had previously not been analyzed to 15 of the 25 STG lines that had already been demonstrated to survive better upon severe drought (Table 1, Figure 1a and Supplementary Figure 1).

To quantify growth of the STG lines, we developed an assay mimicking relatively mild drought stress conditions in which the rosette size of plants grown in soil was followed over time (Material and Methods). To ensure test reproducibility, we also used a large number of plants in an automated platform, designated the 'weighing imaging and watering automated machine' (WIWAM; Figure 1b and Supplementary Figure 2). WIWAM enabled the daily imaging and controlled watering of 216 plants. Plants were germinated and grown under control conditions until stage 1.04 (Boyes et al., 2001), after which watering continued for the control plants, but was stopped for the stressed plants until the set stress level was reached and subsequently kept constant (Supplementary Figures. 3 and 4).

Gene identifier	Gene symbol	Experiment	Line	Biological function	Survival	P-value (genotype control conditions)	P-value (genotype drought conditions)	Mean percent reduction ± SE[a]
At1g01720	ATAF1	1	Loss of function	Transcription	Improved	0.336	0.012	42.35 ± 1.83
At3g06010	cHR12	1	Loss of function	Transcription	Unchanged	0.119	0.724	36.15 ± 2.35
At1g30270	cIpK23	1	Loss of function	Ca+ signaling	Improved	0.21	0.121	42.07 ± 1.87
At1g73660	MApKKK	1	Loss of function	Signaling	Unchanged	0.191	0.118	40.48 ± 3.70
At5g21100	aAAO	1	Loss of function	Reactive oxygen species metabolism	Unchanged	0.836	0.24	42.35 ± 1.83
At5g45340	cYp707A3	1	Loss of function	Hormone metabolism	Improved	0.154	0.657	40.54 ± 2.08
–	–	1	Wild type	–	–	–	–	42.45 ± 2.19
At1g05260	RcI3	2	Gain of function	cell wall	Improved	0.004b	0.027	26.22 ± 6.46
At3g14440	NceD3	2	Gain of function	Hormone metabolism	Improved	<0.001[b]	<0.001[b]	35.98 ± 6.26
At1g78290	SRK2c	2	Gain of function	Signaling	Improved	0.043	0.196	39.38 ± 1.67
At1g74310	HSp101	2	Gain of function	protein stability	Unchanged	0.109	0.102	43.75 ± 6.00
At5g27150	NHX1	2	Gain of function	Transport	Unchanged	0.002[b]	0.209	33.80 ± 2.48
At1g66390	MYB90	2	Gain of function	Transcription	Improved	0.1	0.12	35.83 ± 2.49
At1g77490	tApX	2	Gain of function	Reactive oxygen species metabolism	Improved	0.007[b]	0.001[b]	34.58 ± 2.58
–	–	2	Wild type	–	–	–	–	32.70 ± 3.93
At2g38880	NF-YB	3	Gain of function	Transcription	Improved	0.81	0.858	23.37 ± 1.95
At4g17610	cBL1	3	Gain of function	Ca+ signaling	Improved	0.001[c]	<0.001[c]	27.85 ± 1.69
–	col	3	Wild type	–	–	–	–	29.98 ± 2.70
At5g13680	eLO2	4	Loss of function	Transcription	Improved	<0.001[b]	0.007[b]	35.72 ± 3.64
At3g24500	MBF1c	4	Gain of function	Transcription	Unchanged	0.772	0.29	45.47 ± 2.94
At1g08810	MYB60	4	Loss of function	Transcription	Improved	0.367	0.362	40.03 ± 3.26
At3g15500	ANAc055	4	Gain of function	Transcription	Improved	0.043	0.264	47.75 ± 4.58
At4g09570	cpK4	4	Gain of function	Ca+ signaling	Improved	<0.001[b]	<0.001[b]	47.29 ± 3.71
At1g56600	GOLS2	4	Gain of function	Osmoprotection	Improved	<0.001[b]	<0.001[b]	32.74 ± 2.32
–	–	4	Wild type	–	–	–	–	30.85 ± 5.52
At1g31970	STRS1	5	Loss of function	Transcript stability	Unchanged	0.012	0.026	42.74 ± 5.01
At5g08620	STRS2	5	Loss of function	Transcript stability	Unchanged	0.367	0.316	37.97 ± 3.31
–	–	5	Wild type	–	–	–	–	33.39 ± 1.32
At1g15690	AVp1	6	Gain of function	Transport	Improved	<0.001[c]	<0.001[c]	33.01 ± 1.55
At1g05680	UGT	6	Gain of function	Hormone metabolism	Improved	0.672	0.466	49.91 ± 2.51
–	–	6	Wild type	–	–	–	–	36.89 ± 4.05

Table 1: STG lines tested show no significant genotype-specific responses to the imposed drought stress, either combined or across all the time points. [a]Mean ± SE percent reduction of rosette size by drought measured 10 days into the treatment. [b]Significant ($P < 0.01$) decrease of rosette area of the STG compared with the wild type under control or drought condition. [c]Significant ($P < 0.01$) increase of rosette area of the STG compared with the wild type under control or drought conditions.

In wild-type plants, progressive soil drying resulted in a gradual decrease of growth rates, with a final reduction of the rosette area of 30–40% as a consequence (Figure 1C, Table 1 and Supplementary Figure 4). To assess the performance of the STG lines in terms of genotype differences and genotype-specific responses to the drought stress, we analyzed genotype, environment, time effects and their interactions with a linear mixed model (Table 1, Supplementary Figsures 5 and 6, Material and Methods). Significant ($P < 0.01$) genotype differences were measured for nine lines and in seven cases the difference was significant in both the control and stress environments. The two lines that were significantly larger under control conditions, CBL1 and AVP1, were also larger during drought, and the reverse held true for the smaller lines GOLS2, CPK4, ELO2, NCED3, NHX1, tAPX and RCl3. However, none of the genotypes tested showed a significant specific response to the drought stress imposed either combined or across all the time points (genotype × environment or genotype × environment × time interaction; data not shown). In other words, growth reduction caused by drought was comparable for all genotypes tested. Overall, our data clearly show that enhanced survival under severe drought is not a good indicator for improved growth performance under mild drought conditions. Superior survival under severe drought is often associated with constitutive activation

of water-saving mechanisms, such as stomatal closure, that can, on the contrary, lead to growth penalty (Kasuga et al., 1999).

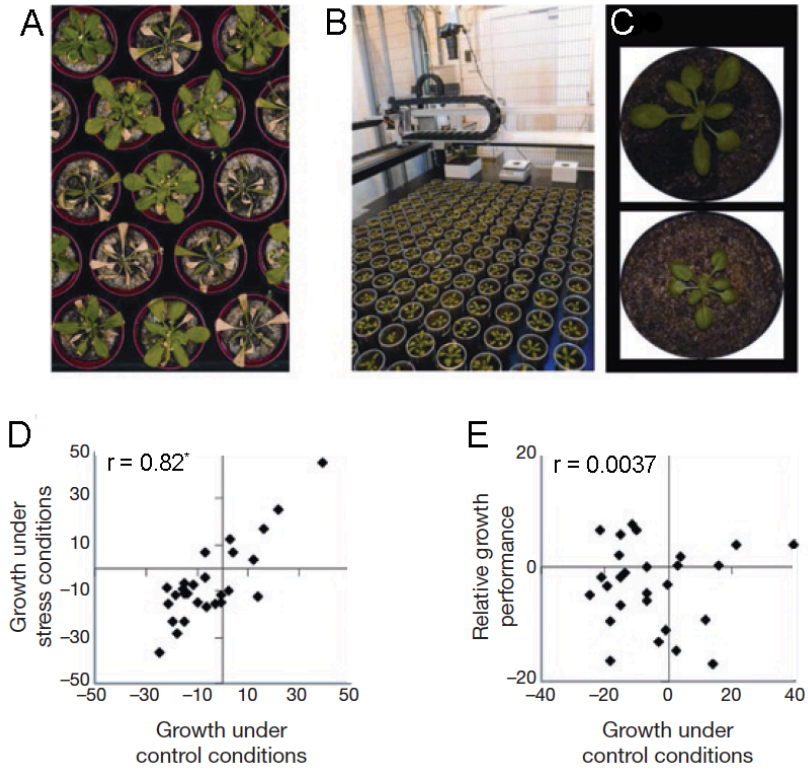


Figure 1: Growth reduction caused by stress is independent of rosette size under control conditions. A,B: Tolerance to severe stress was scored in the survival assay (A), whereas growth under mild drought was assessed with WIWAM in a drought stress regime (B) that reduced final rosette area by 30–40%. C: The top panel shows a wild-type plant grown under control conditions. The bottom panel shows a plant grown under drought conditions. D,E: end-time-point area measurements were used to calculate differences between STG and wild-type (WT) plants under control (area C) (1 – (area C STG/area C WT)) × 100 and drought (area D) conditions (1 – (area D STG/area D WT)) × 100, as well as the difference in drought-related growth inhibition (relative growth performance) ((1 – (area D WT/area C WT)) – (1 – (area D STG/ area C STG))) × 100. Asterisk marks significance ($P < 0.01$).

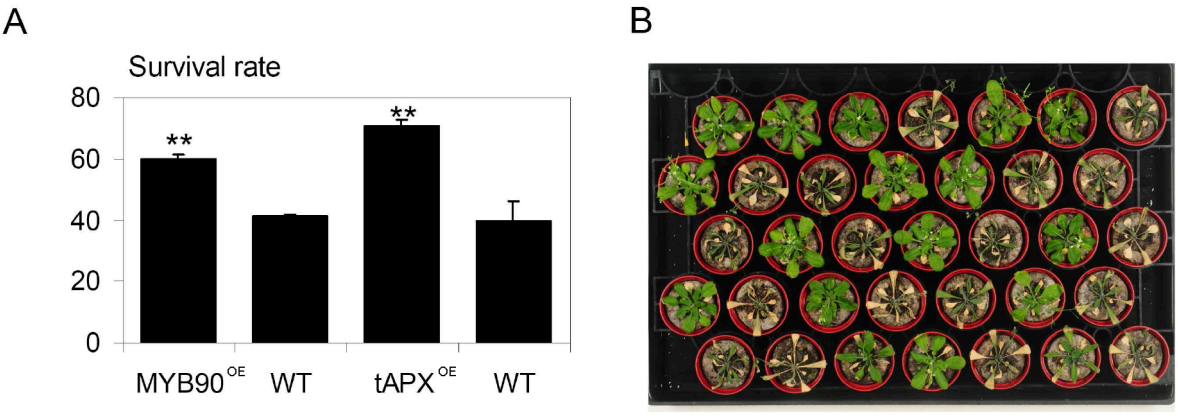
Here, a number of STG lines showed a growth reduction, albeit a subtle one (Table 1). Nevertheless, it is worth mentioning that the observed growth reduction could possibly be bypassed by the use of conditional or tissue-specific promoters (Kasuga et al., 1999; Wang et al., 2009). Generally, plant size and survival, at least under laboratory conditions, are assumed to be negatively correlated, because small plants transpire and use less water. In contrast, in our study, growth reduction caused by mild drought was independent of the STG size measured under control conditions (Figure 1 D and E). Importantly, lines that were larger in the control environment kept their growth advantage under stress conditions, demonstrating that plants have enough resources to sustain both stress tolerance and improved growth. This observation is in line with the favorable carbon status measured in *Arabidopsis* leaves under both mild osmotic and drought stresses (Hummel et al., 2010; Skirycz et al., 2010) and consistent with the hypothesis that plants reduce their growth as a primary adaptation response to stress rather than as a secondary consequence of resource limitation. Under unpredictable environments, growth reduction enables plants to redistribute and save resources, ensuring reproduction even when the stress becomes extreme. However, from the agricultural point of view, when the stress episode does not threaten plant survival, growth reduction can be counterproductive, leading to unnecessary yield loss. Thus, limiting growth reduction might provide a strategy to boost plant biomass productivity under stress.

Biomass yield is of rising importance with the increasing demand for energy crops but has furthermore been shown to be relevant as one component determining seed yield (Hausmann et

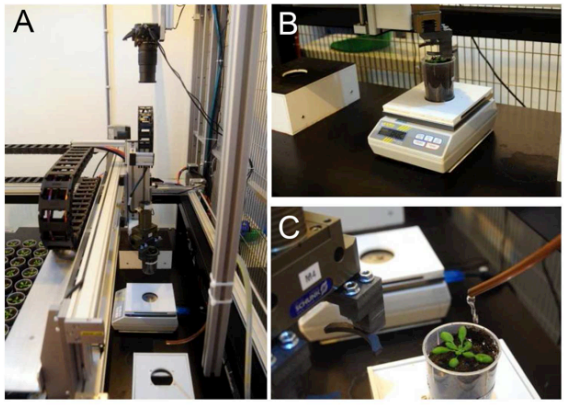
al., 2005). Certainly, a better understanding of the mechanisms that regulate growth under stress conditions, such as those involved in shutting down meristem activity, will be vital in the development of new technologies to increase plant growth under stress (Tisné et al., 2010). Although stress responses of mature organs are relatively well characterized and it is now clear that stress responses are specific to the developmental stage, tissue and even the cell type (Dinneny et al., 2008; Skirycz et al., 2010), the mechanisms that reduced growth under stress are poorly understood. From the technological perspective, automated growth phenotyping under variable environmental conditions with platforms, such as WIWAM, or Phenopsis (Granier et al., 2006) will be essential. Finally, it is important to mention that whereas rosette area can be reliably used to assess biomass, it remains, however, a proxy for seed yield and thus analogous automated platforms focused on seed phenotypes will have to be developed.

In summary, the results imply that indiscriminate selection for lines that survive better under severe stress might be a critical factor responsible for the low success rate by which academic research on drought stress translates to the field. As enhanced survival is largely a function of water-saving mechanisms rather than a net improvement in plant production, it will still be a trait of choice in arid regions, but will most probably not enhance plant yield in moderate climates. In our opinion, such mild, in contrast to severe, conditions will favor bolder plants maintaining more growth, photosynthesis and metabolism despite a water shortage, opening a new exciting paradigm for trait identification.

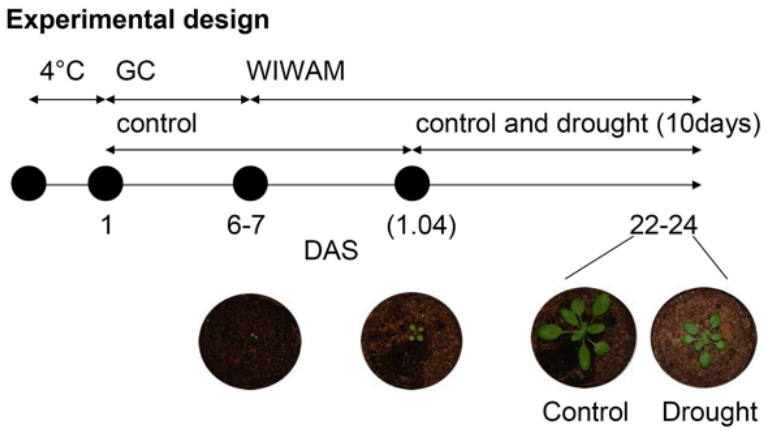
Supplementary figures



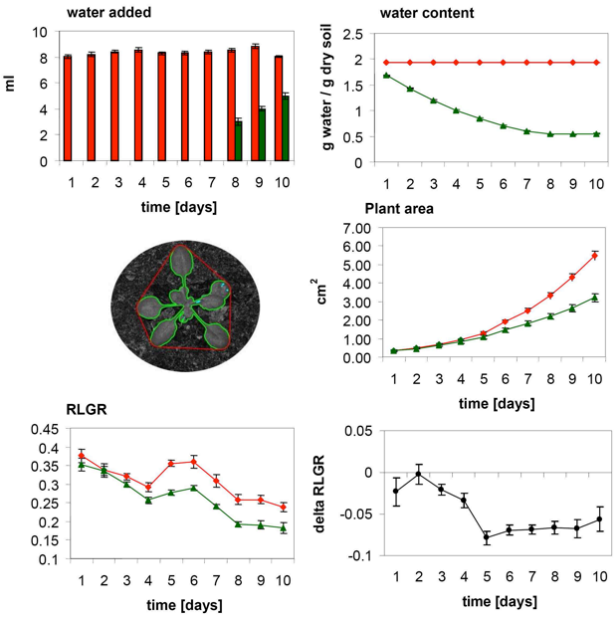
Supplementary Figure 1: Survival assay. A: Survival rate of MYB90 and tAPX over-expressing lines compared to wild-type controls. Data are means ± SE from three independent experiments. Asterisk indicates significance (t-test $P < 0.05$). B: Example of the survival assay 24 h after plants were re-watered.



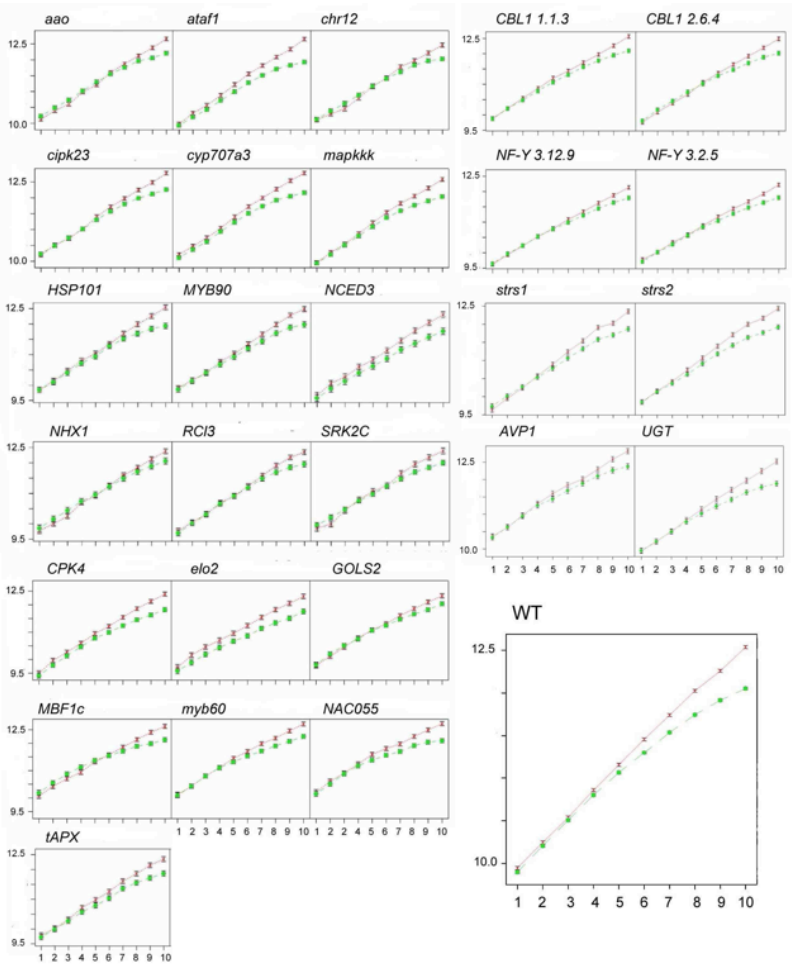
Supplementary Figure 2: WIWAM. A: Imaging, weighing and watering positions. B: Weight measurement. C: Watering.



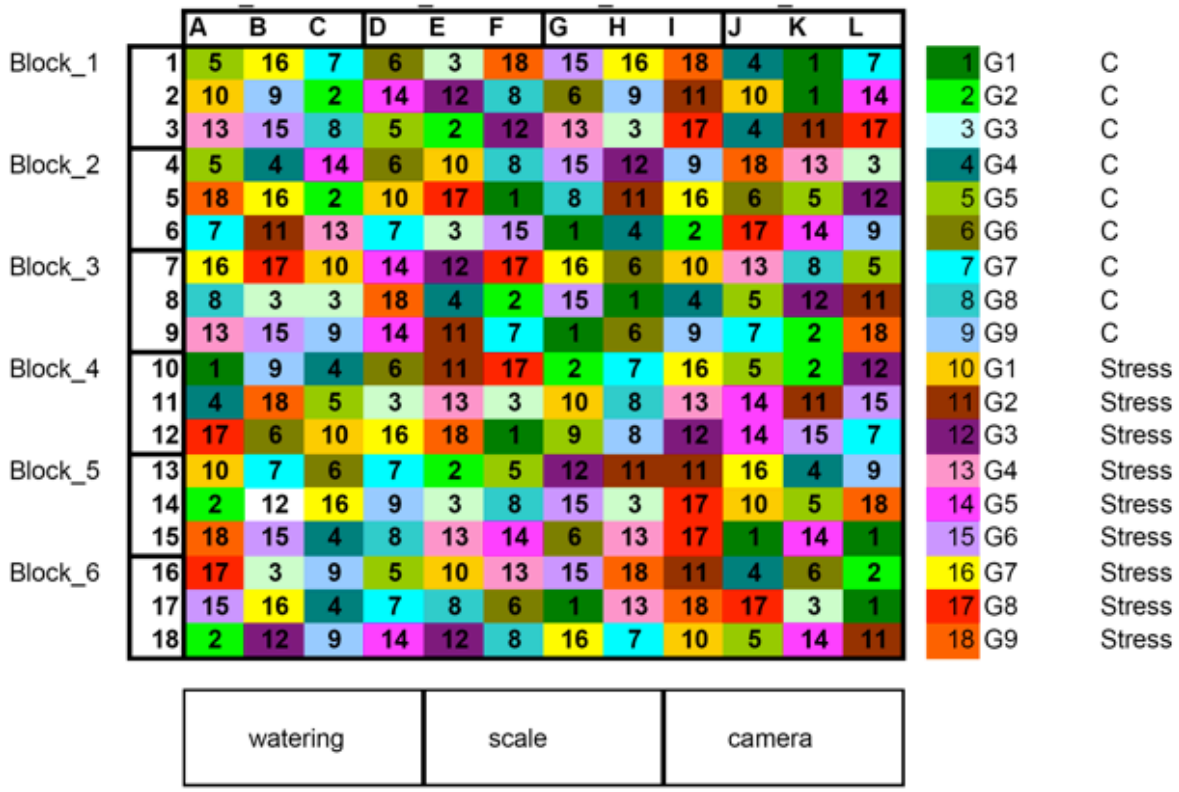
Supplementary Figure 3: Experimental set-up used for drought experiments. Seeds were stratified in cold (4°C) for 2-3 days prior to sowing. Plants were germinated in the long-day (16 h light) growth chamber (GC) and moved to WIWAM at 6-7 days after stratification (DAS). Controlled watering was imposed to all plants until stage 1.04 (4th leaf is approximately 1 mm in size), after which watering continued for control plants, but was stopped for stressed plants until the set stress level had been reached that was kept constant afterward.



Supplementary Figure 4: Example of the output data obtained from WIWAM during 10 days of control (red) and stress (green) treatment. Presented is the amount of water added to control and stress plants calculated from the mass recorded before and after watering and changes in water content over the course of the experiment (calculated based on the mass prior watering). Images were subjected to an automated analysis that provided information on rosette area and perimeter (green line) and hull area (red line). Changes in rosette area can be used to calculate the relative growth rate (RLGR = ln (area day_t) – ln (area day_{t-1}) and the difference (Δ) of RLGR of control and stressed plants.



Supplementary Figure 5: Growth reduction caused by drought compared for all genotypes tested. Means±SE of rosette area (on a ln scale) measured in STG lines (12 independent plants per treatment and from one experiment) and wild-type (WT) plants (120 independent plants per treatment and from six independent experiments) during 10 days of control (red) and drought (green) treatment.



Supplementary Figure 6: Schematic diagram depicting plant randomization on the WIWAM platform. All WIWAM experiments were designed as a complete randomized block design with the genotype × environment combinations occurring exactly twice in each of the six blocks, as depicted on the schematic diagram. Up to nine different genotypes (G1 to G9) were analyzed in control (C) and drought (Stress) conditions that accounted per 12 plants per genotype × environment combination. The platform was divided in six blocks, each block containing two plants per genotype and treatments. Block effects were accounted for in the statistical analysis.

Material and methods

Lines

For overview of the lines, see also Table 1.

Seeds of MBF1c OE were kindly obtained from Ron Mittler (University of Nevada, USA); myb60 and myb90 from Chiara Tonelli (University of Milano, Italy); ANAC055 OE from Vicky Buchanan-Wollaston (University of Warwick, UK); SRK2C OE, GOLS2 OE, NCED3 OE from Kazuo Shinozaki (RIKEN Institute, Japan); CPK4 OE from Da-Peng Zhang (China Agricultural University, Beijing, China); NHX1 OE from Eduardo Blumwald (University of Toronto, Canada); tAPX OE from Irene Murgia (University of Milano, Italy); RCl3 OE from Julio Salinas (University of Madrid, Spain), HSP101 from Susan Lindquist (University of Chicago, USA); strs1 and strs2 from Simon Barack (Ben Gurion University, Israel)n and AVP1 OE from Roberto A. Gaxiola (University of Connecticut, USA). All lines were up–scaled alongside the wild-type plants and expression levels of genes of interest were checked with reverse transcription (RT--PCR or quantitative (q)-RT-PCR. The T-DNA SALK insertion lines for *elo2* (N667190), *chr12* (N605458), *cipk23* (N532341), *mapkkk* (N501982), *aao* (N608854), and *cyp707a3* (N601566) were obtained from NASC. Homozygous and azygous plants were selected by PCR with specific border primers and Lbb1 T-DNA primers. Lack of expression was confirmed with RT-PCR. Open reading

frame of NF-YB and CBL1 were obtained by PCR from the Arabidopsis leaf cDNA library. PCR fragments were cloned with BP reaction into the entry vector and subsequently by LR reaction into the 35S-pXK7 destination vector. Arabidopsis was transformed by floral dip (Clough and Bent, 1998). Single-insertion homozygous plants and azygous wild-type plants were selected by means of kanamycin resistance. Two lines with the highest over-expression level, as determined by Q-RTPCR, were used for phenotypic analysis. The primers used were:

AttB1-NF-YB: GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCGGATACGCCTTCGA

AttB2-NF-YB: GGGGACCACTTTGTACAAGAAAGCTGGGTATTACCAGCTCGGCATTTCTT

AttB1-CBL1: GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGGCTGCTTCCACTCAA

AttB2-CBL1: GGGGACCACTTTGTACAAGAAAGCTGGGTATCATGTGGCAATCTCATCGA.

In-soil plant growth - survival assay

Each plant was grown in a separate pot (55 mm diameter) filled with Jiffy-7 pellets (Jiffy Products, Stange, Norway). Mutant and respective controls were randomized in the same tray (5 x 7 pots), grown under normal conditions (16 h light regime at 21°C) for 14 days before the weight of all pots was equalized. Water was withheld for approximately 2 weeks and plants were re-watered when most of the plants clearly showed symptoms of wilting. Importantly, plants were regularly randomized within the trays during the whole duration of the experiment. Plants that survived were counted and survival was scored. Plants (12 to 18) of each genotype were used to assess survival in three independent experiments.

In-soil plant growth - WIWAM

Plants were grown under a long-day regime (16 h light) at 21°C and 110-120 µmol m^{-2} s^{-1} light intensity. The Weighing Imaging and Watering Automated Machine (WIWAM) is an automated phenotyping platform in which water deficit can be imposed by controlling and stabilizing the soil water status during the development of soil-grown plants. WIWAM was designed together with SMO (Eeklo, Belgium: http://www.smo.be) that also supplied the control electronics. The portal was provided by Automotion (Lovendegem; Belgium: http://www.automotion.be/) and contained linear drives from Bergher Lahr (now Schneider Electric Motion, Freiburg, Germany; http://www.schneider-electric.com). The gripper was purchased from Shunk (Brackenheim-Hausen, Germany; http://www.schunk.com) with a scale from Kern & Sohn (Ballingen, Germany; http://www.kern-sohn.com). WIWAM is controlled by a standard personal computer (PC) with a CAN bus card from Peak Systems (Darmstadt, Germany; http://www.peaksystem.com/). The controlling PC runs a Centeos Linux Operating System that was programmed with in-house developed software written in C and Ruby. This software is available upon request. WIWAM is designed for 216 plants each, grown in cylindrical polypropylene pots (200ml, diameter 53 mm, height 88 mm; VWR International, Leuven, Belgium) that had been perforated to allow fast desiccation. Each pot contains a transponder, linking individual plants to the local database where the information on the watering regime is stored. In a single run, a plant is lifted by an automated robotic arm, moved to the camera position where the transponder is identified, and a photograph is taken; the transponder is recognized, shifted to the scale where dry mass is recorded and used to calculate the amount of

water required, moved to the watering position where water is added, transported back to the scale where the wet mass is recorded, and finally returned to the home position on the platform. The photographs of the plants and the dry and wet masses are automatically stored in the local database. Projected rosette area, perimeter, and compactness (projected rosette area/area of convex hull) are measured automatically by image analysis. Imaging algorithms were written in a Ruby script making use of the Open Source Computer Vision Library (OpenCV) of Intel. Data were stored in a local Mysql database. In the presented experiments, seeds were germinated in 85 g±2 g of Saniflor compost (Van Israel N.V., Geraardsbergen, Belgium; http://www.vanisrael.be) of 71% absolute water content (2.57 water/g dry soil). At 6 or 7 days after stratification (DAS), plants were randomized (Figure S6) and moved to the platform. Controlled watering (set at 68% absolute water content; 2.125 g water/g dry soil) would last until plants reached stage 1.04, after which the stress treatment would start. While control plants were watered daily (to keep the water content at 68%), drought-treated plants were not watered until the water content dropped to 40% (0.66 g water/g of dry soil), which took approximately 7 to 8 days, before the watering would be reinitiated for the next 2-3 days to keep the water content at 40%.

Statistical analysis

Data presented in the current manuscript are from six independent experiments. As all six experiments were designed to have a similar treatment structure, we used the residual maximum likelihood (REML) as implemented in Genstat (version 16) for a combined analysis, also called meta-analysis, of the repeated measurements data. The following linear mixed model was fitted to the data (random terms in bold):

$$y_{ijklmn} = \mu + b_m + g_i + c_j + t_k + gc_{ij} + gt_{ik} + ct_{jk} + gct_{ijk} + \boldsymbol{x_n} + \boldsymbol{e_{ijklmn}}$$

Where y_{ijklmn} is the phenotypic value of the l-th plant from the genotype i measured under condition j at time point k in block m of experiment n, μ is the overall mean term, and e_{ijklmn} is the residual effect; random effects in the model were assumed to be independent and normally distributed with means zero and variance σ_r^2, where r = x (experiment) and e (error). Times of measurements were equally spaced and various ways of modeling the correlation structure (uniform, autoregressive order 1 (AR1) or 2 (AR2), and antedependence order 1 and 2) were compared in the REML framework as implemented in Genstat (Genstat Release 13 Reference Manual, Part 3 Procedure library PL21; VSN International, Oxford). Selection of the best model fit was based on a likelihood ratio test (LRT) statistic and the Aikake Information coefficient (AIC). When residuals from the analysis indicated increasing variance over time, this was modeled directly by specifying that heterogeneity is to be introduced into the model. Significance of the fixed main and interaction effects was assessed by a F-test. Fitting linear contrasts among the levels of factors in the REML analysis of repeated measurements was done with the VTCOMPARISON procedure in Genstat (GenStat Release 14 Reference Manual, Part 3 Procedure Library PL22 (in press); VSN International, Oxford). Here, the cut-off for significance was set to $\alpha = 0.01$ to compensate for the large number of contrasts made.

Acknowledgements

We thank R. Mittler, V. Buchanan-Wollaston, K. Shinozaki, D.-P. Zhang, I. Murgia, J. Salinas, S. Lindquist, E. Blumwald and S. Barack for kindly providing transgenic lines, and M. De Cock for help in preparing the manuscript. This work was supported by grants from Ghent University (Bijzonder Onderzoeksfonds Methusalem project no. BOF08/01M00408) and Multidisciplinary Research Partnership (Biotechnology for a Sustainable Economy no. 01MRB510W), the Interuniversity Attraction Poles Programme (IUAP VI/33), initiated by the Belgian State, Science Policy Office and the European Community Grant FP6 IP AGRON-OMICS (contract LSHG-CT-2006-037704). S.D. is indebted to the Agency for Innovation through Science and Technology for a predoctoral fellowship. V.B.T. is the recipient of a Marie Curie Intra-European Fellowship for Career Development (PIEF-GA-2008-221427).

References

Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Görlach J (2001) Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. Plant Cell 13: 1499–1510

Castiglioni P, Warner D, Bensen RJ, Anstrom DC, Harrison J, Stoecker M, Abad M, Kumar G, Salvador S, D'Ordine R, et al (2008) Bacterial RNA chaperones confer abiotic stress tolerance in plants and improved grain yield in maize under water-limited conditions. Plant Physiol 147: 446–455

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J 16: 735–743

De Block M, Verduyn C, De Brouwer D, Cornelissen M (2005) Poly(ADP-ribose) polymerase in plants affects energy homeostasis, cell death and stress tolerance. Plant J 41: 95–106

Dinneny JR, Long TA, Wang JY, Jung JW, Mace D, Pointer S, Barron C, Brady SM, Schiefelbein J, Benfey PN (2008) Cell identity mediates the response of Arabidopsis roots to abiotic stress. Science 320: 942–945

Granier C, Aguirrezabal L, Chenu K, Cookson SJ, Dauzat M, Hamard P, Thioux J-J, Rolland G, Bouchier-Combaud S, Lebaudy A, et al (2006) PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. New Phytol 169: 623–635

Hausmann NJ, Juenger TE, Sen S, Stowe KA (2005) Quantitative trait loci affecting δ13c and response to differential water availibility in *Arabidopsis thallana*. Evolution 59: 81-96

Hirayama T, Shinozaki K (2010) Research on plant abiotic stress responses in the post-genome era: past, present and future. Plant J 61: 1041–1052

Hummel I, Pantin F, Sulpice R, Piques M, Rolland G, Dauzat M, Christophe A, Pervent M, Bouteille M, Stitt M, et al (2010) Arabidopsis Plants Acclimate to Water Deficit at Low Cost through Changes of Carbon Usage: An Integrated Perspective Using Growth, Metabolite, Enzyme, and Gene Expression Analysis. Plant Physiol 154: 357–372

Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nat Biotechnol 17: 287–291

Nelson DE, Repetti PP, Adams TR, Creelman RA, Wu J, Warner DC, Anstrom DC, Bensen RJ, Castiglioni PP, Donnarummo MG, et al (2007) Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc Natl Acad Sci USA 104: 16450–16455

Skirycz A, De Bodt S, Obata T, De Clercq I, Claeys H, De Rycke R, Andriankaja M, Van Aken O, Van Breusegem F, Fernie AR, et al (2010) Developmental stage specificity and the role of mitochondrial metabolism in the response of Arabidopsis leaves to prolonged mild osmotic stress. Plant Physiol 152: 226–244

Tisné S, SCHMALENBACH I, Reymond M, Dauzat M, Pervent M, Vile D, Granier C (2010) Keep on growing under drought: genetic and developmental bases of the response of rosette area using a recombinant inbred line population. Plant Cell Environ 33: 1875–1887

Umezawa T, Fujita M, Fujita Y, Yamaguchi-Shinozaki K, Shinozaki K (2006) Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. Curr Opin Biotechnol 17: 113–122

Wang Y, Beaith M, Chalifoux M, Ying J, Uchacz T, Sarvas C, Griffiths R, Kuzma M, Wan J, Huang Y (2009) Shoot-specific down-regulation of protein farnesyltransferase (alpha-subunit) for yield protection against drought in canola. Molecular Plant 2: 191–200

2. Leaf responses to mild drought stress in natural variants of Arabidopsis*

Pieter Clauw[1,2], Frederik Coppens[1,2], Kristof De Beuf[3,4], Stijn Dhondt[1,2], Twiggy Van Daele[1,2], Katrien Maleux[1,2], Veronique Storme[1,2], Lieven Clement[3], Nathalie Gonzalez[1,2,#], and Dirk Inzé[1,2,#]

[1]Department of Plant Systems Biology, VIB, Ghent, Belgium
[2]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium
[3]Department of Applied Mathematics Computer Science and Statistics, Ghent University, Ghent, Belgium
[4]Stat-Gent CRESCENDO, Department of Applied Mathematics and Computer Science, Ghent University, Ghent, Belgium

AUTHOR CONTRIBUTIONS

P.C. was the main author of this work and was involved in all practical work and in the biological interpretation of the analyzed data. F.C. analyzed the transcriptome data. K.D.B. and L.C. set up and conducted the differential expression analysis. S.D. developed the image analyses. P.C., T.V.D. and K.M. conducted the experimental work. P.C. and V.S. performed the statistical analysis of the phenotyping data. F.C., N.G. and D.I. supervised the project.

---

* This work was published in Plant Physiology 2015, vol. 167: 800-816.

# These authors contributed equally to this work

Although the response of plants exposed to severe drought stress has been studied extensively, little is known about how plants adapt their growth under mild drought stress conditions. Here, we analyzed the leaf and rosette growth response of six Arabidopsis (*Arabidopsis thaliana*) accessions originating from different geographic regions when exposed to mild drought stress. The automated phenotyping platform WIWAM was used to impose stress early during leaf development, when the third leaf emerges from the shoot apical meristem. Analysis of growth-related phenotypes showed differences in leaf development between the accessions. In all six accessions, mild drought stress reduced both leaf pavement cell area and number without affecting the stomatal index. Genome-wide transcriptome analysis (using RNA sequencing) of early developing leaf tissue identified 354 genes differentially expressed under mild drought stress in the six accessions. Our results indicate the existence of a robust response over different genetic backgrounds to mild drought stress in developing leaves. The processes involved in the overall mild drought stress response comprised abscisic acid signaling, proline metabolism, and cell wall adjustments. In addition to these known severe drought-related responses, 87 genes were found to be specific for the response of young developing leaves to mild drought stress.

Introduction

Since plants started colonizing land, they had to adapt to fluctuations and restrictions in water availability. Morphological adaptations, such as the development of a waxy cuticle, vasculature, stomata, more complex root systems, and a diverse set of molecular mechanisms, allow plants to live in, sometimes, extreme conditions. Numerous studies have elucidated the adaptations of plants to severe drought conditions (McDowell et al., 2008; Akhtar et al., 2012; Golldack et al., 2014), often by withholding water until wilting or by cutting leaves and letting them dry to impose severe water deficits (Iuchi et al., 2001; Llorente et al., 2002; Taji et al., 2002; Cheong et al., 2003, 2007; Tran et al., 2004; Umezawa et al., 2004, 2006; Cominelli et al., 2005; Chen et al., 2006; Nelson et al., 2007; Park et al., 2007; Zhu et al., 2007; Zhang et al., 2008). However, the sudden infliction of such severe drought is unlikely to reflect what naturally happens in the field. In actual field conditions, plants have to adapt continuously to fluctuating environmental parameters, and only rarely will water be present in extreme excess (e.g. flooding) or, conversely, be so low in abundance that it actually threatens plant survival. Frequently, plants experience mild drought stress that, depending on the developmental stage, causes yield losses to various degrees. Despite its potential importance for agriculture, the response of plants to mild drought stress is poorly understood compared with severe dehydration stress (Aguirrezabal et al., 2006; Bouchabke et al., 2008; Harb et al., 2010; Baerenfaller et al., 2012; Des Marais et al., 2012).

Imposing mild drought stress requires a precise and well-monitored experimental setup, including a tight control of the soil water content and defining the precise timing of the drought onset, since the response to stress depends on the developmental stage of the plant (Skirycz et al., 2010; Verelst et al., 2010). Leaf development involves two main processes: cell proliferation and cell expansion. During cell proliferation, corresponding to the first phase of leaf development, all cells are dividing. This phase is followed by cell expansion, starting at the tip of

the leaf and then moving as a front toward the leaf base. The so-called transition phase marks the developmental stage in which fully proliferating leaves convert to leaves with mainly expanding cells (Donnelly et al., 1999; Andriankaja et al., 2012; Gonzalez et al., 2012). Although it is well documented that both cell division and cell expansion are affected by drought stress (Aguirrezabal et al., 2006; Tardieu et al., 2010; Baerenfaller et al., 2012), little is known about the mechanisms involved in the response of the earliest phases of leaf development to mild drought stress.

Arabidopsis (*Arabidopsis thaliana*) is found all around the northern hemisphere, with some small patches along the African coast (Hoffmann, 2002; Koornneef et al., 2004), comprising many different habitats, each with specific environmental conditions, and therefore leading to differences in evolutionary pressure. Since the different populations are genetically isolated due to selfing, it is expected that different Arabidopsis accessions are evolutionarily highly adapted to their local environments. Experiments in laboratory conditions revealed that there is a high plasticity in drought tolerance (Bouchabke et al., 2008), nutrient uptake (Chardon et al., 2010), and salt tolerance (Katori et al., 2010) between different Arabidopsis accessions. With recent advances in sequencing technologies, it is feasible to study the intraspecies variations at the genome and transcriptome level that have risen through the adaptations of the different Arabidopsis accessions to their specific habitats, and also to find what has been evolutionarily conserved.

In order to identify the mechanisms that are active in young, growing leaves exposed to mild drought stress, out of a set of 24 accessions, six accessions capturing most of the variation in drought stress responses were selected (Bouchabke et al., 2008) and subjected to mild drought stress using the automated phenotyping platform WIWAM (Skirycz et al., 2011b), measuring growth-related phenotypes such as rosette area, leaf area, leaf epidermal pavement cell area, cell number, and the stomatal index. It was found that drought affected leaf growth throughout the entire course of development by the interplay of both reduced cell division and expansion and that the accessions behaved differently to mild drought stress.

We also harvested early developing leaf tissue for RNA sequencing and identified a list of 354 genes with common differential expression patterns under mild drought stress in all six accessions. These genes are involved in abscisic acid (ABA) signaling, proline metabolism, and cell wall adjustments. In addition to these known drought-related genes, 87 genes were found to be specific for the response of young developing leaves to mild drought stress. Integration of co-expression and regulatory interaction information showed that the differentially expressed genes are highly connected. Moreover, a few genes were identified as hubs and thus are potential important players in the mild drought stress response in early developing leaf tissue. However, not all differentially expressed genes had a known function or were annotated to a specific process involved in stress responses. Therefore, those genes are interesting candidates to further unravel their role in the mild drought response.

Results

Growth measurements of six Arabidopsis accessions under mild drought

To study the effect of mild drought stress on leaf growth, a protocol was established in which stress is applied to young seedlings using the automated phenotyping platform WIWAM (Skirycz et al., 2011b). This platform allows for the automated weighing, watering, and imaging of the plants and, therefore, strictly controlling the applied watering regime. Because mild drought stress has been shown previously to have a profound effect on the cell number (Aguirrezabal et al., 2006; Pereyra-Irujo et al., 2008; Tardieu et al., 2010; Baerenfaller et al., 2012), drought treatment was started when the third leaf emerges from the shoot apical meristem, at 4 d after stratification (DAS), since at this point during development all cells of the third leaf are dividing. To enable the stress treatment at this early time point, plants were germinated in wet soil and transferred at 4 DAS to mild drought conditions on the WIWAM (Figure 1; see Material and Methods).

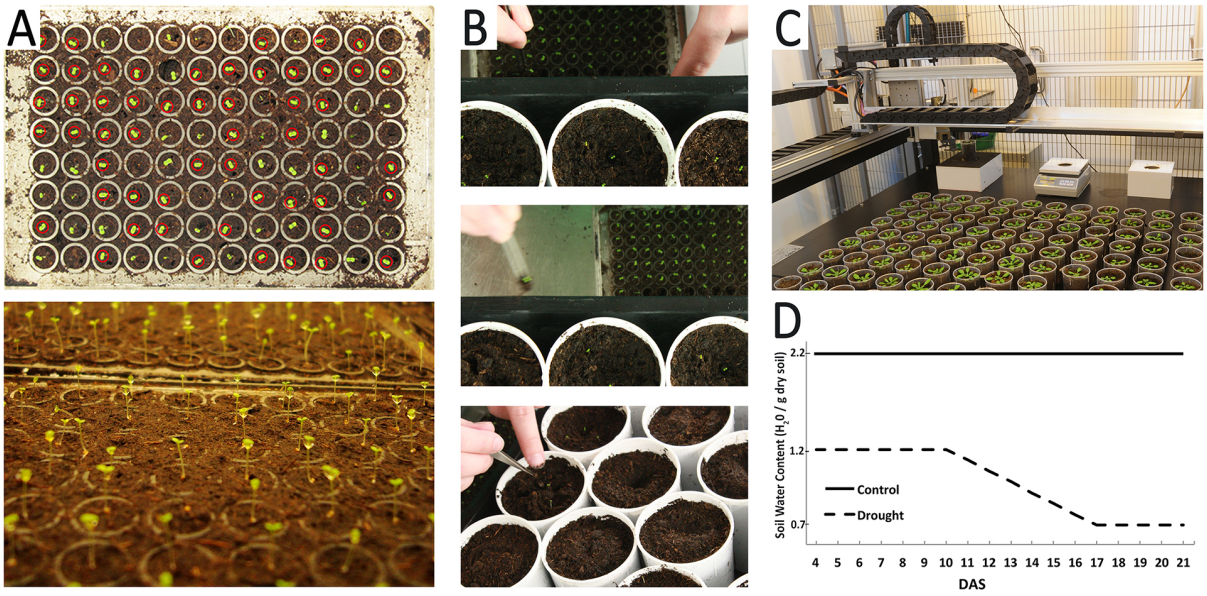


Figure 1: Transfer protocol of seedlings to the WIWAM. A: Seeds were germinated on soil-filled 96-well plates. Average-sized seedlings (54; red circles) were selected using an in-house-developed image analysis algorithm. The seedling selection was done only for phenotyping experiments. B: At 4 DAS, seedlings were manually transferred from 96-well plates to the pretreated soil (drought or control). C: Pots were placed on the WIWAM for automatic phenotyping. D: Control soil water content was maintained at a constant value of 2.2 g water g^{-1} dry soil (solid line) during the entire experiment. For the mild drought condition (dashed line), the soil water content started at 1.2 g water g^{-1} dry soil after transfer to pots at 4 DAS. The stress level increased from 11 DAS onward until it reached 0.7 g water g^{-1} dry soil.

Multiple growth parameters were measured (Supplementary Figure 1): rosette growth was followed over time by calculating the projected rosette area (PRA) daily; the area of the third leaf was measured at the transition from cell proliferation to cell expansion (10–11 DAS) and at maturity (22 DAS); the mean cell number and cell size of the third leaf were determined at maturity together with the stomatal index. Next to these leaf size measurements and cellular analyses, genome-wide transcriptome profiling using RNA-sequencing was performed.

To study the effect of genetic variability in the growth response under mild drought stress applied during the leaf cell proliferation phase, six Arabidopsis accessions (Antwerp-1 [An-1], Bulhary-1 [Blh-1], Columbia-0 [Col-0], Cape Verdi Islands-0 [Cvi-0], Oystese-0 [Oy-0], and Shahdara [Sha]), representing different geographic regions (Supplementary Table 1) and shown to capture most of the variation in drought responses present in a set of 24 accessions (Bouchabke et al., 2008), were used.

Rosette growth in response to mild drought

The dynamic nature of rosette growth was grasped by following the PRA over time with daily intervals for each individual plant grown on the automated imaging platform WIWAM. To identify significant differences in growth between the accessions, a mixed model was used to analyze measurements of PRA over time (see Material and Methods). This statistical model, using time, genotype, and treatment as fixed factors, showed that there was a significant difference in PRA between the accessions' responses to mild drought stress over time (genotype × treatment × time interaction; $P = 0.01$).

Next, we determined which accessions were significantly different by performing a pairwise comparison. This comparison was done for each time point: for the PRA in control conditions, for the PRA under mild drought stress, and for the relative difference of PRA under mild drought compared with control conditions. Results of the pairwise comparisons for all phenotypes can be found in Supplementary Table 2. Under control conditions, over time, Oy-0 was the largest and Cvi-0 was the smallest of the six accessions ($P < 0.05$; Figure 2A). At maturity (21 DAS), Oy-0 reached a PRA of 633 mm^2, while Cvi-0 measured 239 mm^2 (Figure 2, A and B), both differing significantly ($P < 0.05$) from the PRAs of An-1, Blh-1, Col-0, and Sha, which were not significantly different from each other. After exposure to mild drought conditions, from 14 DAS onward, significant reductions in PRA were detected (Figure 3). Therefore, the response over time to the imposed mild drought stress was analyzed from that time point onward by performing pairwise comparisons, showing that Oy-0 and Cvi-0 differed significantly from Blh-1 and Sha in their growth response over time to mild drought stress ($P < 0.05$; Figure 3). At maturity (21 DAS), Oy-0 and Cvi-0 were 72% and 67% smaller, respectively, under mild drought compared with control conditions (Figure 2B). Whereas the response of the PRA over time was different for Cvi-0 (Figure 3), at the final time point the reduction in PRA of Cvi-0 did not differ from An-1, Blh-1, Col-0, and Sha, in contrast to that of Oy-0, which was significantly different from all other accessions except Cvi-0 ($P < 0.05$).

In conclusion, our results demonstrate that mild drought has different effects on leaf growth in different accessions.

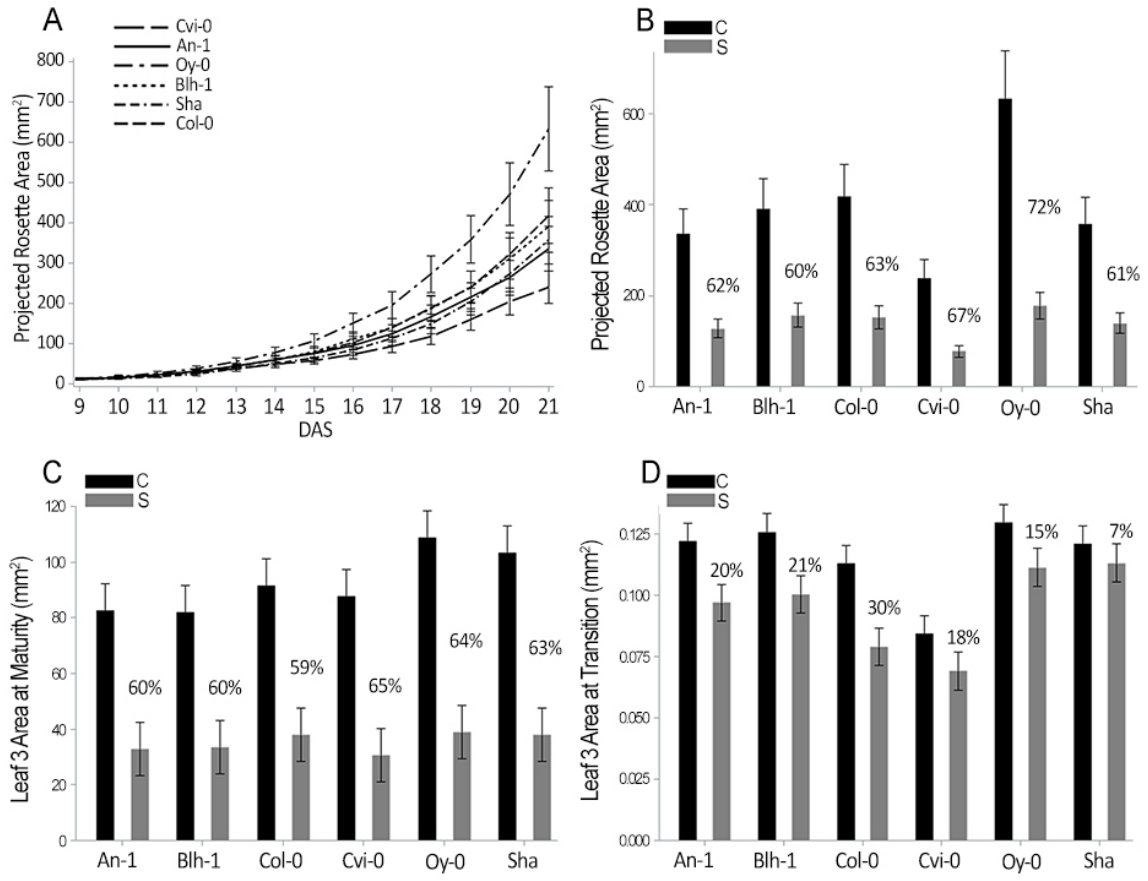


Figure 2: Measurements of rosette area and leaf size. A: Projected rosette area over time in control conditions for the six accessions. B: Projected rosette areas at maturity. Control conditions (C) are indicated in black, and mild drought conditions (S) are indicated in gray. C: Leaf 3 areas at maturity. D: Leaf 3 areas at the transition from proliferation to expansion. All values are least-square means ± SE estimated from the mixed model, and percentages represent reductions under mild drought relative to the control.

Leaf growth in response to mild drought

In order to analyze the cellular nature of the differences in growth response to mild drought stress, we studied in detail the third leaf harvested from the six accessions grown under control and stress conditions and quantified the leaf area, cell number, cell size, and stomatal index of the abaxial epidermis.

Under control conditions, the mature third leaf area ranged from 82 mm^2 in Blh-1 to 109 mm^2 in Oy-0 (Figure 2C). The two accessions with the largest third leaf, Oy-0 and Sha (103 mm^2), did not differ significantly from each other, but both had a significantly larger third leaf than the other four accessions ($P < 0.05$). The leaf area decreased for all six accessions when exposed to mild drought stress, ranging from 31 mm^2 in Cvi-0 to 39 mm^2 in Oy-0 ($P < 0.05$). Upon mild drought treatment, no significant differences in leaf area were detected between the accessions. The accessions showing the highest reduction at maturity were Cvi-0 (65%), Oy-0 (64%), and Sha (63%; Figure 2C).

In order to know to what extent the drought treatment affected the third leaf early during development, leaf size was measured at the transition from proliferation to expansion. Upon

microscopic investigation for the presence of puzzle-shaped pavement cells near the leaf tip, we concluded that for Cvi-0 the transition started at 11 DAS, whereas for all other accessions this was at 10 DAS. At this transition, the areas of leaves grown in control conditions ranged from 0.084 mm^2 in Cvi-0 to 0.130 mm^2 in Oy-0 (Figure 2D). After mild drought treatment, Oy-0 and Sha (0.111 and 0.113 mm^2, respectively) showed the largest third leaf, whereas that of Cvi-0 measured 0.069 mm^2 and therefore was the smallest of the six accessions (Figure 2D).

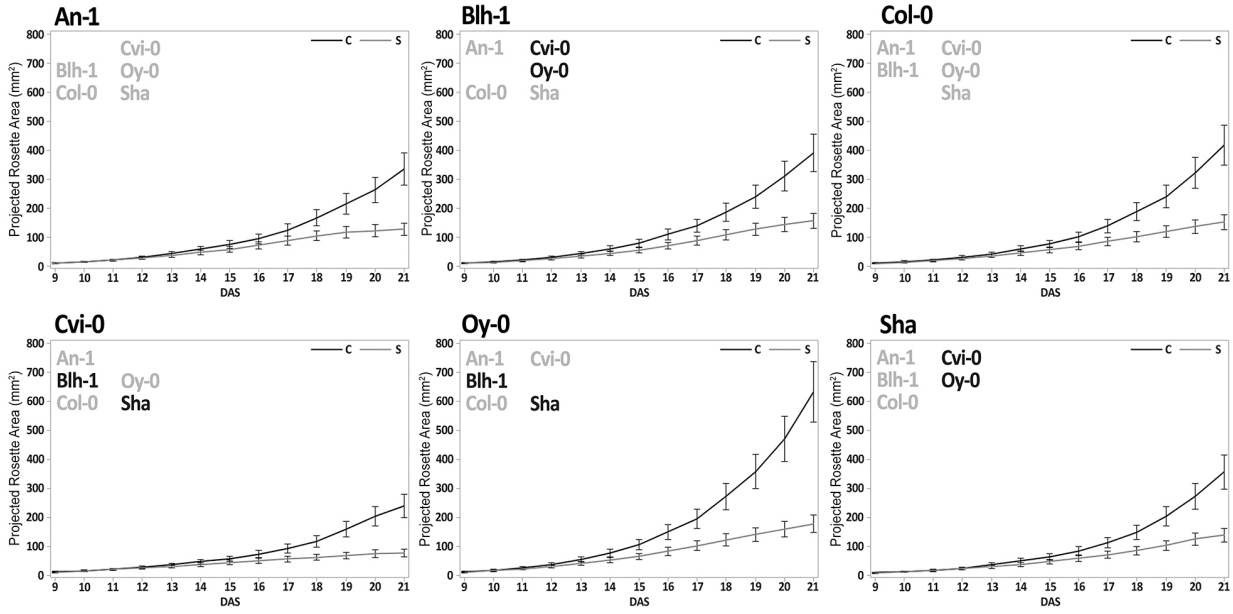


Figure 3: Projected rosette area over time under control (C) and mild drought (S) conditions for the six accessions. Accession names indicated in the figure in black are significantly different from the reference accession in the time × genotype × treatment interaction at a P value cutoff of 0.05, and those indicated in gray are not significant. Values are least-square means ± SE estimated from the mixed model.

The reductions in size under mild drought stress relative to the control condition varied from 7% in Sha to 30% in Col-0 (Figure 2D). Whereas the 7% size reduction in Sha was not found to be significant, the leaf area of the other accessions was reduced significantly under mild drought stress at this early developmental time point. In order to elucidate the cellular characteristics of the changes in leaf size, cellular drawings were made of the epidermis of the third leaf at maturity. These drawings were analyzed using an in-house-developed algorithm (Andriankaja et al., 2012) to obtain the cell number, cell area, and stomatal index. Under mild drought stress, we found that both pavement cell area and number were significantly reduced in all six accessions ($P < 0.05$). The reduction for pavement cell area (Figure 4A) was quite similar for all accessions, ranging from 43% (Col-0) to 54% (Sha and Oy-0), in contrast to the number of pavement cells, which differed significantly more between accessions and varied from 18% (Blh-1 and Oy-0) up to 41% (An-1; Figure 4B). The reduction in cell size and leaf size led to a higher cell density (Figure 4C); however, the stomatal index was not significantly different between control and mild drought conditions (Figure 4D).

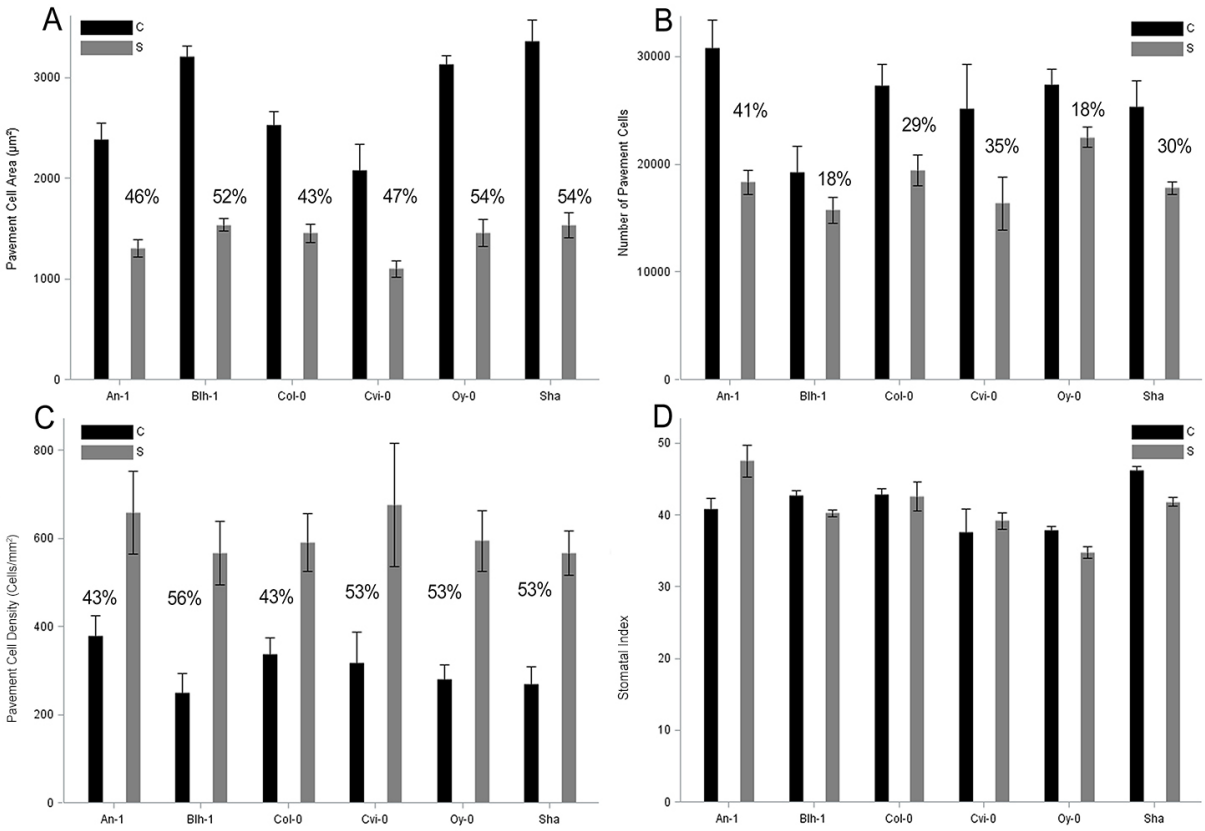


Figure 4: Measurements of cellular parameters of the mature third leaf of six accessions in control and mild drought conditions. A: Averaged cell areas of the pavement cells (mm²). B: Total number of pavement cells in the leaf. C: Pavement cell density (cells mm^{-2}). D: Stomatal index as a percentage of stomata on total cell number. Control conditions (C) are indicated in black, and mild drought conditions (S) are indicated in gray. All values are means ± SE, and percentages represent reductions under mild drought relative to the control.

In conclusion, the size of the third leaf was reduced in all six accessions after mild drought treatment, although to a different extent, due to decreases in cell number and average cell size.

Transcriptome analysis

Previous work has shown that the transition from cell proliferation to cell expansion is of pivotal importance in determining the final leaf size (for review, see Gonzalez et al., 2012). This study aims to further unravel the effect of mild drought stress on this size-determining phase of leaf development. In our experiments, the third leaf was harvested at the beginning of the transition from cell proliferation to cell expansion (10–11 DAS), when the first expanding cells are discerned at the tip of the leaf. RNA was extracted and gene expression was determined through RNA sequencing.

Gene expression under mild drought stress

Common and accession-specific differentially expressed genes upon mild drought treatment were prioritized using a stage-wise statistical analysis (for details, see Material and Methods; Supplementary Figure 2). In the first stage, the differential expression was assessed for each gene. This stage consisted of two tests: stage I.a (for accession) checked for differential expression in at least one of the six accessions, resulting in 265 genes; and stage I.c (for common) checked for

differential expression on average over the six accessions, giving 359 genes. The union of both tests delivered 439 genes that were differentially expressed during mild drought (Supplementary Table 3).

In order to find genes with an accession-specific response to mild drought stress, stage II tested the genes with significant differential expression in at least one accession (from stage I.a) for accession specificity (i.e. accession × treatment interaction). The stage II test identified 60 (Supplementary Table 3) out of the 265 stage I.a genes to show accession-specific differential expression, referred to as accession-specific genes. However, further analysis with pairwise tests assessing for differences in mild drought stress induced by differential expression between accessions in stage III (see Material and Methods) unveiled that none of the 60 genes showed a response to the mild drought stress in one accession that was significantly different from the response in all five remaining accessions. The expression profiles of the 60 accession-specific genes in the six accessions for the three biological repeats are shown in Supplementary Figure 3.

In the next step, the 359 genes from the stage I.c analysis were analyzed further. For statistical reasons (detailed in "Material and Methods and Supplementary Figure 2), five genes were removed. The remaining 354 genes are referred to as common drought genes (Supplementary Table 3). The expression profiles of all 354 genes in the six accessions for the three biological repeats are shown in Supplementary Figure 4.

A Gene Ontology (GO) enrichment analysis was performed to gain insight into the functional categories of the 354 common drought genes (Maere et al., 2005; Table 1; Supplementary Table 4). The top enriched categories involve various abiotic stress responses, such as ABA signaling, osmotic stress, reactive oxygen species, and salt stress. Besides stress-related GO categories, cell wall modification and cell growth-related genes also were clearly enriched among the common drought genes.

Description	Q-value	Number of Genes
cellular response to hormone stimulus	1.39E-05	15
plant-type cell wall loosening	1.39E-05	7
abscisic acid mediated signaling pathway	1.55E-05	8
cell wall modification	3.32E-05	11
hormone-mediated signaling pathway	4.65E-05	14
cell wall organization	5.35E-05	12
plant-type cell wall modification	6.25E-05	7
response to osmotic stress	9.38E-05	18
plant-type cell wall organization	1.89E-04	8
response to biotic stimulus	7.16E-04	20
cell wall organization or biogenesis	7.67E-04	13
regulation of transcription, DNA-dependent	5.77E-03	23
response to oxidative stress	6.14E-03	11
response to salt stress	1.18E-02	13
cellular response to reactive oxygen species	1.22E-02	3
regulation of gene-specific transcription	1.27E-02	3
cell wall thickening	1.54E-02	3
developmental growth involved in morphogenesis	1.76E-02	8

regulation of primary metabolic process	1.88E-02	34
ethylene mediated signaling pathway	2.11E-02	4
cellular response to ethylene stimulus	2.25E-02	4
glucan metabolic process	2.25E-02	6
cell growth	2.70E-02	9
response to cold	2.75E-02	9
developmental growth	3.05E-02	8
cell wall modification involved in multidimensional cell growth	3.43E-02	3
response to ethylene stimulus	3.43E-02	6
proline metabolic process	3.44E-02	2
response to temperature stimulus	3.92E-02	11

Table 1: Main enriched GO categories in the 354 common drought genes. The full list of enriched GO categories can be found in Supplementary Table 4. Q-value is the p-value of the enrichment corrected with the Bonferroni method for multiple testing.

Hormone signaling

The imposed mild drought stress had a major impact on ABA signaling in young developing leaves. Six of the 14 members of the PYRABACTIN RESISTANCE (PYR)/ PYRABACTIN RESISTANCE1-LIKE (PYL)/ REGULATORY COMPONENTS OF ABA RECEPTORS (RCAR) ABA receptor protein family (PYR1, PYL1, PYL4, PYL5, PYL6, and PYL8) were down-regulated, whereas six of the nine protein phosphatase 2Cs (ABSCISIC ACID INSENSITIVE1 [ABI1], ABI2, HIGHLY ABSCISIC ACID INDUCED PP2C GENE1 [HAI1], HAI2, PROTEIN PHOSPHATASE 2CA [PP2CA/AHG3], and HYPERSENSITIVE TO ABSCISIC ACID1 [HAB1]), two of the nine ABA responsive element-binding factors (ABFs; ABF2/ ABSCISIC ACID-RESPONSIVE ELEMENT-BINDING PROTEIN1 [AREB1] and ABF3), and six of the ABF target genes (ABSCISIC ACID-INSENSITIVE FIVE BINDING PROTEIN1 [AFP1], PP2CA/AHG3, HAI1, 3-KETOACYL-COA SYNTHASE2 [KCS2], RESPONSIVE TO DESSICATION20 [RD20], and AT5G53390; Yoshida et al., 2010, 2014) were up-regulated. In addition, 37 genes previously associated with ABA (Nemhauser et al., 2006; The Arabidopsis Information Resource; Supplementary Table 5) were found to be differentially expressed. For example, D1-pyrroline-5-carboxylate synthase (P5CS1), a rate-limiting enzyme in the proline biosynthesis pathway (Strizhov et al., 1997), was up-regulated, whereas the gene encoding proline dehydrogenase (PROLINE DEHYDROGENASE1 [PDH1]/EARLY RESPONSIVE TO DEHYDRATION5 [ERD5]/PROLINE OXIDASE [POX]) was down-regulated. Taken together, 57 of the 354 differentially expressed genes were ABA related.

Ethylene was previously associated with the regulation of leaf growth *in vitro* under mannitol-mediated growth reduction (Skirycz et al., 2011a; Dubois et al., 2013). Under mild drought conditions in soil, two 1-aminocyclopropane-1-carboxylic acid (ACC) oxidases (ACO4 and ACO2) were down-regulated as well as four ethylene response factors (ERFs; ERF2, ERF15, AT5G07580, and AT5G61590).

DELLAs are known negative regulators of growth under various stresses (Achard et al., 2006, 2008; Magome et al., 2008; Navarro et al., 2008; Claeys et al., 2012). Therefore, we compared the common differentially expressed genes upon mild stress with a list of putative DELLA targets

(Claeys and Inzé, 2013). Of the 354 common drought genes, 43 were putative DELLA targets (Supplementary Table 6), which was a highly significant enrichment (P = 2.67E-45).

Cell wall modifications

At least 21 common drought genes were involved in cell wall modifications. The cell wall-loosening expansins (EXPA1, EXPA3, EXPA4, EXPA15, EXPB1, and EXPB3), pectin lyases (AT1G10640, AT1G60590, AT1G67750, AT3G61490, AT4G13710, and AT4G24780; Cosgrove, 2005), and pectin methylesterase inhibitors (AT1G23205 and AT2G26440) were up-regulated. On the other hand, genes encoding cell wall-strengthening enzymes, xyloglucan endotransglucosylases/hydrolases (XTHs) and fasciclin-like arabinogalactans (FLAs; Cosgrove, 2005; MacMillan et al., 2010), were consistently down-regulated (FLA2, FLA9, MERISTEM5 [MERI5B], XTH6, XTH9, XTH15, and XTH16).

Co-expression network and regulatory interactions

The mild drought stress imposed on the six accessions clearly provoked a differential regulation of genes that were previously associated with drought stress, such as ABA signaling genes. However, of the 354 common differentially expressed genes, 216 were previously not associated with drought-responsive mechanisms (based on drought-related GO terms, relation to ABA, and involvement in cell wall modifications; Supplementary Table 7), of which 37 have yet unknown functions.

Next, we used the online tool CORNET (De Bodt et al., 2010, 2012) to perform co-expression and regulatory interaction network analyses. All 354 common differentially expressed genes were tested against different predefined microarray data sets in CORNET (see Material and Methods) and resulted in a network of 202 co-expressed genes. In addition, a regulatory network analysis was performed in CORNET to get a view on the regulatory interactions between the co-expressed genes. This analysis used the confirmed regulatory interactions from the AGRIS (Davuluri et al., 2003; Palaniswamy et al., 2006; Yilmaz et al., 2011) and microarray gene-target relation databases. In addition, the text-mining database EVEX (Van Landeghem et al., 2013) was used to find extra regulatory interactions described in the literature. Interactions with neighbor genes were included for the regulatory interactions, adding 144 genes to the network, bringing the total to 346 genes. The co-expression and regulatory interaction analyses showed high connectivity between the differentially expressed genes (Figure 5).

The 20 genes with the highest number of interactions (both co-expression and regulatory) are given in Supplementary Table 8. The genes with the most interactions encode *MITOGEN-ACTIVATED PROTEIN KINASE 3* (*MPK3*; 37 interactions) and a protein of unknown function (AT4G36500; 28 interactions). Also, a jasmonic acid (JA)-synthesizing lipoxygenase (*LOX2*; 17 interactions) is among the 20 most interacting genes.

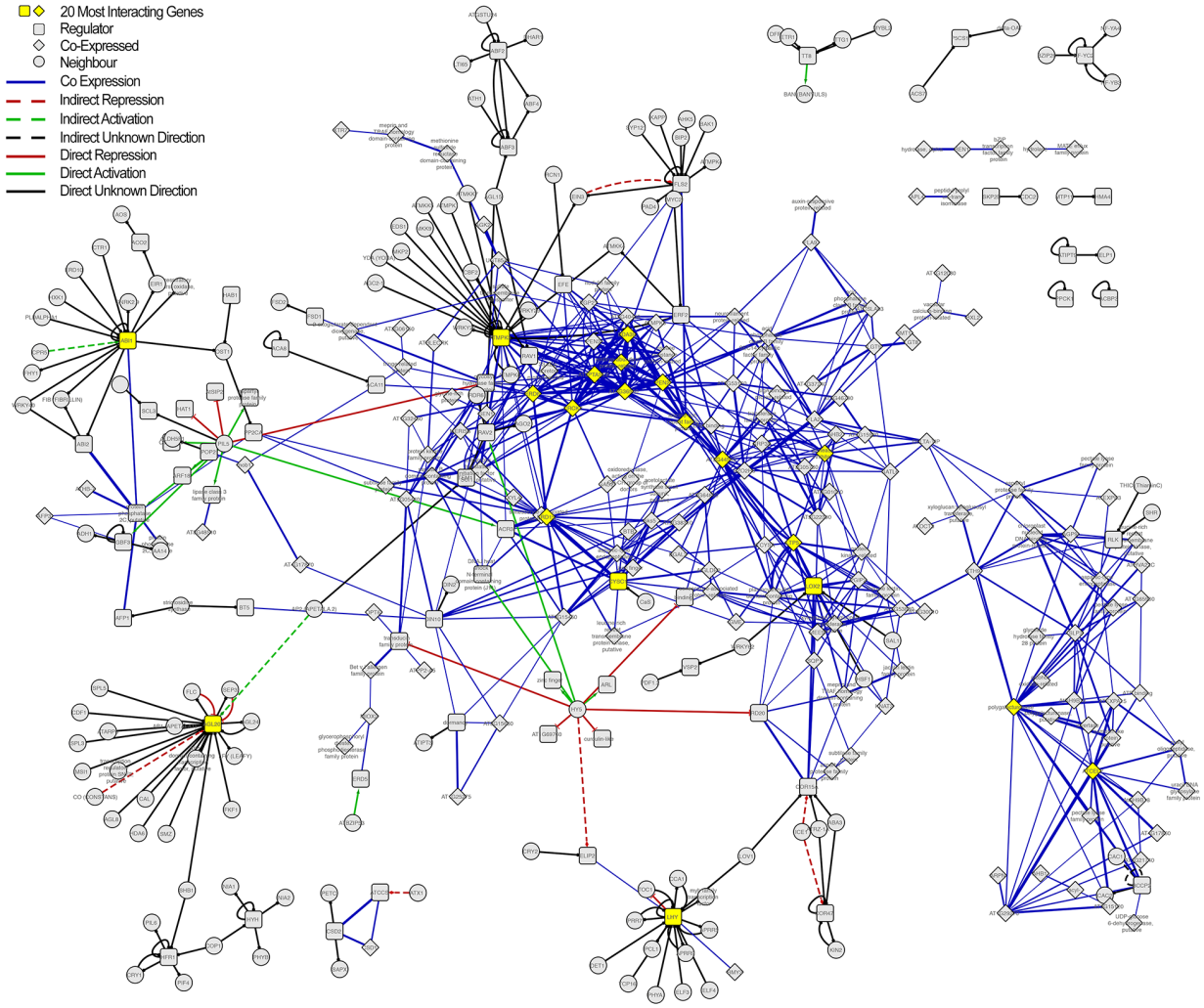


Figure 5: Co-expression and regulatory interaction network of common differentially expressed genes in the six accessions. Edges colored blue connect co-expressed genes, and thickness increases with rising co-expression coefficients. Red, green, and black lines represent regulatory interactions, which can be direct (solid lines) or indirect (dashed lines). Regulatory interactions can be activating (green), repressing (red), or unknown (black). The top 20 most interacting genes in the network are indicated in yellow and can be found in Supplementary Table 8. Squares and diamonds are query genes, and circles are neighbor genes. Visualization is based on the co-expression analysis done in CORNET with a co-expression coefficient of 0.7 and confirmed regulatory interactions from AGRIS, microarray gene-target relations, and EVEX.

Comparison with severe drought studies

In the past, a large number of studies focused on severe stress. In order to make a comparison with our mild drought study, three studies imposing severe progressive drought stress on soil-grown plants were selected (Huang et al., 2008; Matsui et al., 2008; Harb et al., 2010). Whereas Harb et al. (2010) used Affymetrix ATH1 micro arrays (approximately 21,000 genes), the two other studies applied custom-made arrays (approximately 30,000 genes) with probes for all 354 common drought genes except two. A total of 188 of the common drought genes were differentially expressed in the same direction in at least one of the three studies compared (Supplementary Figure 5). Forty-two genes were part of the core ABA signaling machinery (12 genes) or responded to ABA (30 genes), including the Pro-biosynthesizing P5CS1 and proline dehydrogenase (PDH1/ERD5/POX). Eight from the 20 cell wall-modifying genes, including all four XTHs and the XTH-similar MERI5B, found in mild drought, behaved similarly in severe

drought. Interestingly, four out of five differentially expressed expansins in mild drought showed opposing expression in severe drought. From the 43 putative DELLA targets differentially expressed in mild drought, 25 showed a similar behavior to severe drought.

In the end, 166 common drought genes were not transcriptionally modified in severe drought imposed on mature leaf tissue. Subsequently, we analyzed the behavior of the common drought genes in mature tissue exposed to mild drought stress.

Comparison with mild drought in mature tissue

Only three studies have addressed the transcriptomic response of soil-grown Arabidopsis plants to mild drought stress (Harb et al., 2010; Baerenfaller et al., 2012; Des Marais et al., 2012). In all cases, the transcriptional responses of mature tissues were analyzed, with the exception of the study by Baerenfaller et al. (2012), which pooled different developmental stages to detect mild drought stress-responsive genes. A comparison of the 354 common drought genes described in this study with the above-listed experiments showed that 216 of the 354 common drought genes in young developing tissue were expressed in the same direction in mature tissue under mild drought (Supplementary Figure 6). Except for one protein phosphatase type 2C (PP2C; HAB1) and ABF2, all ABA signaling genes overlapped with the studies in mature tissue. Also, 21 ABA-responsive genes behaved similarly in mature and young developing leaf tissue. Furthermore, 15 of the 20 cell wall-modifying common drought genes and 12 of the 43 putative DELLA targets also played putative roles in mature leaves exposed to mild drought stress.

Genes specifically expressed in young developing tissue subjected to mild drought stress

By comparing with the above-listed studies, 87 genes were identified to be differentially expressed specifically in young developing tissue under mild drought stress conditions (Figure 6; Supplementary Table 9). Strikingly, none of these genes, with the exception of *MYO-INOSITOL POLYPHOSPHATE 5-PHOSPHATASE 2*, *COLD-REGULATED 47*, *MPK3*, *NACL-INDUCIBLE GENE 1*, and AT5G53390, were previously found to be associated with ABA. Five of the 87 common drought genes were involved in cell wall modification: EXPA15, three genes encoding pectin lyases (AT1G60590, AT1G67750, and AT3G61490), and one pectin methylesterase inhibitor (AT1G23205).

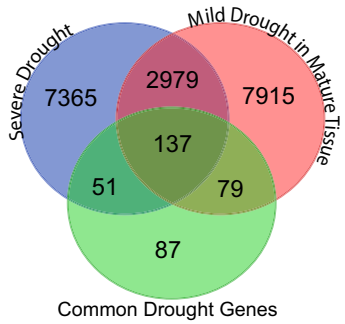


Figure 6: Overlap between genes involved in severe drought studies (Huang et al., 2008; Matsui et al., 2008; Harb et al., 2010), genes in mature tissue in mild drought studies (Harb et al., 2010; Baerenfaller et al., 2012; Des Marais et al., 2012), and the common drought genes.

Many of the 87 genes have not been associated previously with known growth responses in mild drought. However, some of the common drought genes might play a role in regulating growth in mild drought in young developing leaf tissue; interesting candidates are *MPK3* and *CYTOCHROME P450 78A7* (CYP78A7).

Discussion

Mild drought scenarios, their effects and robustness

Most of our knowledge about the responses of plants to drought stress is obtained by studying severe dehydration stress. The physiological relevance of this type of experiments compared with field conditions, however, has been questioned (Verslues et al., 2006; Claeys and Inzé, 2013; Lawlor, 2013). Moreover, increased tolerance to severe drought often comes with a yield penalty (Yang et al., 2010) and does not result in sustained growth under mild drought conditions (Skirycz et al., 2011b; Claeys and Inzé, 2013). The aim of this study was to analyze the effect of mild drought stress on young developing leaves. To this end, the automated phenotyping platform WIWAM was used to impose stress early during leaf development when the third leaf emerges from the shoot apical meristem. Daily imaging allowed for the analysis of stress effects in six different accessions on rosette growth over time. Although the six accessions analyzed responded differently to the imposed stress at the phenotypic level, the molecular response, investigated by RNA sequencing of developing leaves at the transition from proliferation to expansion, revealed that similar stress-related processes were differentially regulated in all accessions.

The protocol described here allows for an early onset of the mild drought stress when the third leaf emerges from the shoot apical meristem. In previous studies, stress onset started at the emergence of leaf 6 (Aguirrezabal et al., 2006; Bouchabke et al., 2008), explaining some major differences observed in the sensitivity of accessions to mild drought stress. For example, the effect of mild drought stress in An-1 previously was found to be rather limited (Aguirrezabal et al., 2006; Bouchabke et al., 2008), whereas in this study, a considerable size reduction of the third leaf and PRA was found. Differences in the developmental timing at which the drought stress was applied and in the drought scenario could lead to different outcomes in terms of drought tolerance or sensitivity (Tardieu, 2012). Therefore, it is important to screen different drought scenarios in order to come to a robust consensus response, which might be of great interest for agricultural applications. To increase the robustness of this consensus response, it is favorable to test plants with different genetic backgrounds.

Cellular response to mild drought stress

Mild drought stress in all six accessions caused reductions in cell area and cell number. Pavement cell area was reduced by 42% to 55%, depending on the accession, and cell number was decreased by 16% to 24%, except for An-1, having a 41% reduction. Reductions in cell number and cell size by drought have been reported previously (Aguirrezabal et al., 2006; Pereyra-Irujo et al., 2008; Baerenfaller et al., 2012), and quantitative trait loci linked to this reduction have been described (Tisné et al., 2008). In our experimental setup, cell expansion was relatively more affected by the mild drought stress than cell proliferation. Although the mild drought stress

started and was effective during the proliferation phase, the water content was further lowered during the expansion phase from 11 DAS onward (Figure 1D). This increase in drought severity during the transition phase from proliferation to expansion can explain the larger effect on cell expansion compared with that on cell proliferation. Alternatively, an initial reduction in cell division could have been compensated for during leaf development by meristemoid divisions, generating extra pavement cells while forming stomata (Geisler et al., 2000; Bergmann and Sack, 2007). Compensation by meristemoid division has been observed previously in the leaf growth response to mannitol (Skirycz et al., 2011a).

Transcriptome response to mild drought in six accessions

In total, 354 genes were found to be differentially expressed similarly in all six accessions under mild drought. Despite the phenotypic differences between the accessions, which often can be explained by differences in regulation of the same genes (Chen et al., 2005; Delker et al., 2010), our analysis revealed few differences in the transcriptome response to the mild drought stress between the accessions. Only 60 genes, with functions in seemingly unrelated pathways, were found to have a drought response that differs between the accessions. None of these 60 genes showed a differential expression response that was specific for one accession only.

This small number of accession-specific genes was in contrast to a previous report (Des Marais et al., 2012) showing that, in mature leaves exposed to mild drought stress, there are many more genes accession-specifically expressed. A possible explanation of this discrepancy is that our experimental setup with RNA sequencing of microdissected young developing leaves introduced a certain amount of variation between biological repeats. This additional variation lowered the power for finding interactions and significant pairwise differences. Furthermore, it is also plausible that the transcriptional response to mild drought conditions in young developing leaves, containing mainly dividing cells, was more conserved than the response observed in mature leaves.

Therefore, we focused on the genes that were found to be common and differentially expressed under mild drought stress in the six accessions. These genes were involved in ABA signaling, proline metabolism, and cell wall metabolism.

ABA plays a key role in the response of soil-grown plants to mild drought stress

The role of ABA in severe drought stress applied to mature plants is well known (Yamaguchi-Shinozaki and Shinozaki, 2006; Finkelstein, 2013). Also in the mild drought stress setup described here, numerous ABA signaling genes were found to be differentially expressed in early developing leaves exposed to mild drought stress. Among the ABA signaling genes were genes encoding ABA receptors (PYR/PYL/RCAR proteins), PP2Cs, and ABFs. The PYR/PYL/RCAR protein family contains 14 members that are all capable of activating signaling in response to ABA. Most members have been described as ABA receptors that, in the presence of ABA, inhibit PP2C through protein binding (for review, see Cutler et al., 2010; Umezawa et al., 2010). This, in turn, allows the accumulation of phosphorylated SUCROSE NONFERMENTING1-RELATED PROTEIN KINASE2s (SnRK2s) and subsequent

phosphorylation of ABFs (Wang et al., 2013), which were shown to interact with DEHYDRATION-RESPONSIVE ELEMENT-BINDING PROTEIN (DREB)/C-REPEAT BINDING PROTEINs such as DREB1A, DREB2A, and DREB2C, playing a central role in the drought response (Cutler et al., 2010; Lee et al., 2010). Mild drought stress imposed during early leaf development down-regulated the expression of six ABA receptor-encoding genes (PYR1, PYL1, PYL4, PYL5, PYL6, and PYL8). PYR1 and PYL1 are able to inhibit PP2Cs only in the presence of ABA, whereas PYL4, PYL5, PYL6, and PYL8 can inhibit PP2Cs in an ABA-independent way (Miyakawa et al., 2013). The absence of differentially expressed ABA biosynthesis genes such as 9-CIS-EPOXYCAROTENOID DIOXYGENASE (Tan et al., 2003), together with the down-regulation of the receptors and up-regulation of six PP2Cs (ABI1, ABI2, HAI1, HAI2, PP2CA/AHG3, and HAB1) as inhibitors of ABA signaling, might reflect a negative feedback on the ABA response. Yet, two of the four main ABFs (ABF2/AREB1 and ABF3; Yoshida et al., 2014) and six of the ABF target genes (AFP1, PP2CA/AHG3, HAI1, KCS2, RD20, and AT5G53390; Yoshida et al., 2010, 2014) were up-regulated, suggesting that the ABA response is still active despite such a putative negative feedback. In addition to these six ABF target genes, 37 other ABA-induced genes were differentially expressed in our data set, confirming the importance of ABA in the transcriptome response to mild drought stress also in proliferating leaves. The SnRK2 genes were not found to be differentially expressed, which is not surprising, since their activity is regulated through phosphorylation (Finkelstein, 2013).

One of the biochemical pathways controlled by ABA is proline metabolism. Proline acts as an osmolyte, but it also functions as osmoprotectant and reactive oxygen species detoxifier. In addition, proline has been suggested to have a regulatory role in mitochondrial functioning, cell proliferation, cell death, and the expression of rehydration-inducible genes (for review, see Szabados and Savouré, 2010). In our data, we found that the proline biosynthesis gene P5CS1 was up-regulated, whereas the catabolic proline dehydrogenase-encoding gene (PDH1/ERD5/POX) was down-regulated. This is in agreement with previous observations of increased proline levels in response to drought stress (Verslues and Sharma, 2010).

Ethylene is one of the hormones involved in the biotic stress response and was found to play a role in the growth inhibitory response to low concentrations of mannitol administered to in vitro-grown plants (Verelst et al., 2010; Dubois et al., 2013). This treatment was generally thought to trigger osmotic stress, but the recent identification of specific mannitol receptors changed this view. Mannitol is now thought to trigger a biotic stress response besides its function as osmoticum (Claeys et al., 2014b; Trontin et al., 2014). Nevertheless, after mannitol treatment, the transcription factor ERF6 triggers, in an ACC-dependent manner, the expression of GA2-OXIDASE6 (GA2-OX6). Increased expression of GA2-OX6 leads to GA inactivation and reduced GA levels, in turn stabilizing the DELLA proteins, which will inhibit leaf growth (Dubois et al., 2013). The transcriptome analysis of developing leaves exposed to mild drought stress provided no evidence for the up-regulation of the ethylene signaling pathway. On the contrary, down-regulation of the expression of the genes encoding ACC oxidases (ACO2 and ACO4), one putative ACC oxidase (AT2G25450), and four ERFs (ERF2, ERF15, AT5G07580, and AT5G61590) was observed. Similar results were found in leaf tissue by Baerenfaller and colleagues (2012), who found that genes encoding two ACC oxidases (ACO2 and ACO5) and two ACC synthases (ACS10 and ACS11) were down-regulated, whereas two other genes

encoding ACC synthases (ACS8 and ACS12) were up-regulated upon drought stress. However, as in the mannitol response, there was a clear growth reduction. Thus, other growth-regulating mechanisms must be at work during mild drought stress.

The general down-regulation of ethylene biosynthesis and signaling, and the suggested active role of ABA in the mild drought response, are in agreement with a proposed antagonism between both hormones. ABA and ethylene show contrasting effects on a number of processes, such as stomatal aperture (Tanaka et al., 2005), hyponastic growth (Benschop et al., 2005), seed germination (Beaudoin et al., 2000; Ghassemian et al., 2000), and defense gene expression and disease resistance (Anderson et al., 2004; De Paepe et al., 2004). Both hormones are suggested to antagonistically regulate each other's metabolism and signaling (Cheng et al., 2009). Our data suggest that this also might be the case in response to mild drought stress.

DELLA proteins are known to play a role in stress-induced mitotic exit and cell differentiation (for review, see Claeys et al., 2014a). DELLA proteins are degraded in the presence of GA and are important players in GA signaling. In total, 43 putative DELLA targets were differentially expressed in all six accessions exposed to mild drought stress. DELLAs thus seem to play a role in the growth response of transitioning tissue under mild drought. Besides ethylene, GA and ABA also are known antagonists in different processes (Gómez-Cadenas et al., 2001; Ye and Zhang, 2012; Lü et al., 2014). Under mild drought, ABA might inhibit the GA signaling pathway, leading to DELLA stabilization and thus a reduction in growth. Further research is required to elucidate the molecular mechanism of this antagonism under mild drought stress.

Cell wall adjustments

In drought conditions, cell walls of organs and tissues that need to maintain their growth are loosened to keep them in a growth-ready state, while other tissues stiffen their cell walls (Wu and Cosgrove, 2000; Moore et al., 2008). In total, 21 genes involved in cell wall modifications were differentially expressed in young proliferating leaves exposed to mild drought stress. The up-regulated genes encode the cell wall-loosening expansins, pectin lyases, and pectin methylesterase inhibitors. On the other hand, genes encoding XTHs and fasciclin-like arabinogalactans, both involved in cell wall strengthening (Cosgrove, 2005; MacMillan et al., 2010), were down-regulated. The finding that mild drought stress results in the up-regulation of genes involved in cell wall loosening and the down-regulation of genes encoding cell wall-strengthening enzymes is an indication that, in this condition, the growing leaf tissue is kept in a growth-ready state. Moreover, since the sampled tissue was transitioning from proliferation to expansion, it is likely that under mild drought the cells are preparing for expansion earlier in comparison with control conditions, resulting in a differential up-regulation of cell wall-loosening enzymes. The earlier preparation for expansion, and thus the earlier arrest of cell proliferation, might be the driving force for the reduction in cell number observed under mild drought. Up-regulation of cell wall-loosening enzymes under mild drought also allows for growth with less turgor pressure and might facilitate cell expansion after rewatering (Lechner et al., 2008).

Central players in the drought-regulated network and new drought-responsive genes

Besides known stress-related genes, more than half of the differentially expressed genes have no previously described function in drought responses. However, co-expression analysis revealed that a majority of the genes were co-expressed in a large set of microarray data sets. A further examination of the co-expression network allowed for the identification of genes connected to up to 37 other genes either by co-expression or regulatory interactions.

The most connected gene or hub in the obtained network encoded MPK3 and could potentially be an important regulator in the mild drought response. MPK3 has been shown to be involved in the regulation of the biotic stress response (Nakagami et al., 2005; Pedley and Martin, 2005; Beckers et al., 2009). In both transcriptional (through WRKY33) and posttranscriptional ways, MPK3 is able to induce and maintain ethylene biosynthesis (Li et al., 2012). Furthermore, MPK3/MPK6 is capable of phosphorylating ERF6 (Meng et al., 2013), a central player in the ethylene-regulated growth response to stress (Dubois et al., 2013). Under mild drought, we found that the expression of MPK3 was down-regulated specifically in young developing tissues in all six accessions, in agreement with the observed down-regulation of ethylene biosynthesis and signaling. In addition, MPK3 has been suggested to be a mediator of the MITOGEN-ACTIVATED PROTEIN KINASE KINASE4-induced osmotic stress response (Kim et al., 2011). In conclusion, our data suggest that MPK3 could play a central role in the growth response to mild drought stress.

Another highly connected gene in the network was LOX2, which is required for JA synthesis (Bell et al., 1995), a hormone involved in pathogen defense, wounding response, and plant growth (Zhang and Turner, 2008). The reduction in plant growth upon JA treatment is caused by an arrest at the G2 phase of the cell cycle (Swiątek et al., 2002, 2004; Pauwels et al., 2008). Recently, LOX2 was shown to be under antagonistic control of TEOSINTE BRANCHED1/CYCLOIDEA/ PROLIFERATING CELL FACTOR 4 (TCP4) and TCP20. TCP4, similar to the rest of the class II TCPs, is regulated by JAGGED AND WAVY (Palatnik et al., 2003) and is known to cause a decrease in cell proliferation resulting in smaller leaves when hyperactivated (Sarvepalli and Nath, 2011). TCP20 (a class I TCP), on the other hand, is suggested to stimulate the cell cycle and organ growth through binding to the regulatory sequences of *CYCLIN B1;1, PROLIFERATING CELL NUCLEAR ANTIGEN*, and ribosomal genes (Li et al., 2005). The class I and class II TCP transcription factors are proposed to regulate leaf development through JA as an intermediary hormonal signal. According to this model, JA levels would increase during leaf development and, as such, induce the transition from proliferation to expansion (Danisman et al., 2012). Thus, the up-regulation of LOX2 under mild drought might lead to increased JA concentrations and to an earlier transition to cell expansion compared with control conditions. This faster transition is in agreement with the response of the genes encoding cell wall modifiers. A faster transition might explain the reduced cell number and, thus, part of the size reduction induced by mild drought stress. The importance of JA for the growth reduction is strengthened by the observation that JA-insensitive mutants (coronatine-insensitive1 and jamsonate-insensitive1) lack growth reduction under mild drought (Harb et al., 2010).

Strikingly, the second and eighth most interacting genes in the co-expression and regulatory interaction network are of unknown function (AT4G36500 and AT5G44580) and are interesting candidates for further functional characterization.

The specific response of young developing leaf tissue to mild drought stress

With a large number of studies focusing on severe drought and the awareness that tolerance to severe drought does not necessarily mean better growth under mild drought (Skirycz et al., 2011b), the question remains how specific the response of plants to these two different manifestations of water deficit is.

A bit more than half of the common drought genes found here to be transcriptionally altered by mild drought also were previously reported to be affected by severe drought. Many of the genes affected by both mild and severe drought have a role in ABA perception and signaling. Also, Harb et al. (2010) found a strong overlap for ABA-related genes between mild and severe drought in mature leaves. Here, we show that ABA also plays a role in the response of young developing leaves to mild drought. A striking difference between the response of plants to mild or severe drought stress, however, is the behavior of genes encoding cell wall-modifying proteins. Whereas genes encoding XTHs showed the same expression behavior in severe and mild drought stress, genes encoding expansins were, opposite to mild drought stress, down-regulated in severe drought. A plausible explanation is that in severe drought, no growth can be allowed, whereas in mild drought conditions, the expansins allow cell growth under reduced turgor pressure.

Of particular interest are the 87 genes specific for young developing leaf tissue exposed to mild drought. Twelve of the 87 genes have no known function and are interesting candidates for further functional analysis. Another 10 genes encode proteins that are involved in cell wall synthesis, loosening, or remodeling: four pectin lyases (AT1G60590, AT1G67750, AT3G61490, and AT4G13710), two cellulose synthase-like proteins (CSLB03 and CSLC6), one expansin (EXPA15), one extensin (EXT3), one arabinogalactan protein (AGP9), and one pectin methylesterase inhibitor (AT1G23205). All these genes are up-regulated specifically in young developing tissue under mild drought. Although growth is reduced, loosening the cell wall allows a reduced growth under lower turgor evoked by the mild drought.

Three genes encoding transcription factors, a Myb-like protein (AT5G56840), *BASIC HELIX-LOOP-HELIX 42*, and *TEMPRANILLO 1* (*TEM1*), are found among the genes that are differently expressed in young developing leaves exposed to mild drought. TEM1 has been shown to directly repress GA3-OX1 and GA3-OX2 expression (Osnato et al., 2012), both involved in one of the main bioactive GA-forming steps. Here, a down-regulation of TEM1 was noticed, which would allow an increase in GA, which might antagonize the observed ABA response. Other potential regulators specific for developing leaves are MPK3 and CYP78A7. As detailed above, MPK3 is the most central player in the co-expression and regulatory interactions network of mild drought stress-responsive genes.

CYP78A7 and CYP78A5/KLUH encode cytochrome P450 proteins and have redundant roles in positively regulating the relative growth of the shoot apical meristem and developing leaves and

flowers (Anastasiou et al., 2007; Wang et al., 2008; Eriksson et al., 2010). The observed up-regulation of CYP78A7 is seemingly in disagreement with the coinciding growth reduction. However, to maintain growth under mild drought at reduced levels, there probably is a complex network of both growth promoters and growth inhibitors forcing each other into a certain balance.

In conclusion, the use of different accessions allowed for the detection of a robust set of genes that play a role in the mild drought response in different genetic backgrounds. Whereas some of these genes also were found to be altered in response to severe drought stress or in the response to mild drought in mature tissue, a unique set of 87 genes was found to specifically play a role in the response to mild drought in young developing leaves.

MATERIAL AND METHODS

Plant growth and experimental setup

Six Arabidopsis (*Arabidopsis thaliana*) accessions (An-1, Blh-1, Col-0, Cvi-0, Oy-0, and Sha) were grown in a growth chamber under controlled environmental conditions (21°C, 55% relative humidity, 16-h day/8-h night, and 110– 120 mmol m^{-2} s^{-1} light intensity). Seeds of Blh-1 were provided by Olivier Loudet (Institut National de la Recherche Agronomique). Other accessions were obtained from the Nottingham Arabidopsis Stock Centre (N22660). The decision to screen these six accessions was based on an article by Bouchabke et al. (2008), where these six accessions covered a large part of the variation in response to mild drought present in their collection of 24 accessions mainly for total leaf area, but electrolyte leakage, relative water content, and cut rosette water loss also were taken into account. All accessions were bulked at the same time. The geographical origin of the six accessions is provided in Supplementary Table 1. Plants were germinated on soil-filled 96-well plates after 4 d of stratification at 4°C in the dark. The timing of transition was checked for each accession; therefore, Cvi-0 was put to germinate 1 d earlier, as it reached the transition phase of the third leaf 1 d later compared with the other accessions. Seedlings were transferred to pots at 4 DAS. Fifty-four average-sized seedlings were selected (Figure 1A) using an in-house-developed image algorithm in order to reduce seedling size effects at the beginning of the experiment. Before transfer, the relative water content of the pots was set at 1.2 g water g^{-1} dry soil for the mild drought treatment; the control condition was set at 2.2 g water g^{-1} dry soil (Figure 1D). Once the seedlings were transferred to pots (Figure 1B), they were placed on the automated phenotyping platform WIWAM (Skirycz et al., 2011b). The water content of the soil was kept constant until 10 DAS, after which it was lowered daily to target 0.7 g water g^{-1} dry soil for the mild-drought-treated plants (Figure 1D). Images of the rosette of each plant were taken daily until 21 DAS and analyzed for the PRA (Skirycz et al., 2011b). Size measurements of the third leaf were done at the transition from cell proliferation to cell expansion (10–11 DAS) and at maturity (22 DAS). For practical reasons, the mature third leaf was harvested 1 day later than the last PRA measurement at 21 DAS. To this end, the leaves were cut from the rosette, cleared in ethanol, and transferred to lactic acid before mounting on microscope slides. Measurements based on microscope images were done using ImageJ

(http://imagej.nih.gov/ij/), and analysis of drawings made from the abaxial epidermis allowed for quantification of the cell area, cell number, and stomatal index (Andriankaja et al., 2012).

Statistical analysis of phenotypic measurements

All statistical analyses were performed with SAS/STAT software, version 9.4 of the SAS System for Windows 7 64bit. (SAS Institute; http://www.sas. com). A mixed model was fitted to the leaf 3 area data, both at transition and maturity. The fixed effects part of the model contained the main effects genotype and treatment and their interaction. The replication factor experiment was included in the model as a random factor to account for the correlation between observations belonging to the same experiment. Type 3 F tests were performed to estimate the pairwise comparisons between the accessions at control and stress levels as well as the pairwise comparisons between the accessions of the decrease in leaf 3 area due to stress. P values were adjusted for multiple comparisons with the false discovery rate (FDR) implemented in the SAS multtest procedure. Effects were considered significant at an FDR threshold of 0.05. Cellular measurements were analyzed like the leaf 3 area data. In the case of a significant interaction (at the 0.05 significance level) between genotype and treatment, simple F tests of effects were performed to estimate differences between the stress and control treatments for each genotype. In the absence of a significant interaction but the presence of significant main effects, all pairwise comparisons were calculated between the genotypes and between the treatments. P values were adjusted with Tukey's method as implemented in SAS. A longitudinal data analysis was conducted in SAS on the daily measured PRA. A mixed model was used including the main effects of genotype, treatment, time, and all two-way and three-way interaction terms. The PRA was log transformed to stabilize the variance. Various covariance structures were tested to model the correlations between measurements done on the same plant. The Toeplitz covariance structure was selected based on Akaike's information criterion. The replication factor experiment was included in the model as a random factor. The Kenward-Rogers approximation for computing the denominator degrees of freedom for the tests of fixed effects was used (Schaalje et al., 2001). The appropriate contrasts were defined in the estimate statement to conduct pairwise comparisons for the different factors and final time point. Type 3 F tests were used to test significant differences. P values were adjusted for multiple comparisons with the FDR. An effect was considered significant at an FDR threshold of 0.05. Residual diagnostics were carefully examined.

Transcriptome analysis

Sampling

To ensure sufficient material for transcriptome analysis, 60 seedlings were grown per accession per treatment. Plants were harvested at 10 DAS except for Cvi-0, which was harvested at 11 DAS. Plants were flash frozen in liquid nitrogen immediately upon harvest. RNAlater-ICE (Ambion), which prevents RNA from degrading, was cooled at 270°C, added to the samples, and allowed to penetrate the tissue at 220°C for 5 d. The third leaf was collected by microdissection with a microscope. Samples were microdissected in a petri dish on dry ice to keep the samples

below room temperature. Dissected leaves were ground with a Retsch machine and 3-mm metal balls. Samples were obtained from three independent biological repeats.

RNA extraction

RNA was extracted with Trizol (Invitrogen) according to the manufacturer's instructions. RNA samples were subjected to DNA digestion with the RNase-free DNase I kit (Qiagen), and, subsequently, impurities were removed with the RNeasy mini kit (Qiagen).

RNA-sequencing

Library preparation was done using the TruSeq RNA Sample Preparation Kit version 2 (Illumina). In brief, poly(A)-containing mRNA molecules are reverse transcribed, double-stranded complementary DNA is generated, and adapters are ligated. After quality control using the 2100 Bioanalyzer (Agilent), clusters are generated through amplification using the TruSeq PE Cluster Kit v3-cBot-HS (Illumina) followed by sequencing on the Illumina HiSeq2000 with the TruSeq SBS Kit v3-HS (Illumina). Sequencing was performed in paired-end mode with a read length of 100 bp. The quality of the raw data was verified with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; version 0.9.1). Next, quality filtering was performed using the FASTX-Toolkit (http://hannonlab.cshl. du/fastx_toolkit/; version 0.0.13): reads were globally filtered in which, for at least 75% of the reads, the quality exceeds Q20, and 39 trimming was performed to remove bases with a quality below Q10, ensuring a minimum length of 90 bp remaining. Repair was performed using a custom Perl script. Reads were subsequently mapped to the Arabidopsis reference genome (The Arabidopsis Information Resource 10) using GSNAP version 2012-07-20 (Wu and Nacu, 2010), allowing maximally five mismatches. The concordantly paired reads that uniquely map to the genome were used for quantification at the gene level with htseq-count from the HTSeq.py python package (Anders et al., 2014). Data were normalized using TMM, implemented in the edgeR package (version 3.5.27) in R (version 3.1.0; Robinson et al., 2010; McCarthy et al., 2012).

Differential expression analysis

The samples were harvested in four different batches. The transcriptomes for both control and mild drought stress were observed for each accession within a batch. Gene-wise RNA sequencing counts were analyzed using a negative binomial model with batch, accession, and treatment main effects, an accession × treatment interaction, and a tag-wise overdispersion parameter for each gene. We aimed to discover genes for which the drought stress response is accession specific as well as genes with a common drought-induced differential expression pattern in the different accessions. The former genes can be identified by testing for changes in stress-induced differential expression between accessions. Our design, however, implies 15 pairwise differential expression comparisons for each gene. It is well known that high-throughput experiments with a huge number of simultaneous hypothesis tests typically have a low power for detecting significant multiple treatment effects (Jiang and Doerge, 2006). Similar to Jiang and Doerge (2006), we propose a stepwise procedure for increasing the power to detect accession-specific differentially expressed genes between control and mild drought. In stage I, the null hypothesis of no differential expression between control and mild drought is tested against the alternative hypothesis that there is differential expression in at least one accession (test I.a). The

list with significant I.a genes will be enriched for genes with a drought stress-induced expression response that is accession specific. Significant I.a genes are tested for a drought stress × accession interaction in stage II. Finally, stage II genes with a significant interaction are further dissected in stage III by assessing all multiple pairwise comparisons of mild drought stress-induced differential expression between the different accessions. Genes with a common differential expression between mild drought and control can be prioritized by assessing the average differential expression between mild drought stress and control over all accessions. This involves only one contrast for each gene and thus can be assessed using a single-stage hypothesis test, which is referred to as test I.c and will complement test I.a in stage I. Five of the genes detected in test I.c also appeared in the stage II list of genes with accession × treatment interaction and were removed from the downstream analysis. We follow the same rationale as Jiang and Doerge (2006) to control the joint FDR over all three stages of the proposed test procedure. They showed that the overall FDR in a stepwise multiple comparison procedure can be controlled below the prespecified FDR significance level a when the sum of the FDR significance levels used in each of the different stages is below a. They also showed that it is advantageous to choose a_1 in the first stage larger than the remaining stages, so that more genes can enter stage II. Here, the overall FDR over tests I.a and I.c in the first stage is controlled at 4%, while the FDR in stage II and III is set at 0.5%. All analyses were conducted with the edgeR package (version 3.5.27) in R (version 3.1.0; Robinson et al., 2010; McCarthy et al., 2012).

GO enrichment

GO enrichment analyses were conducted with BiNGO (Maere et al., 2005).

Co-expression analysis

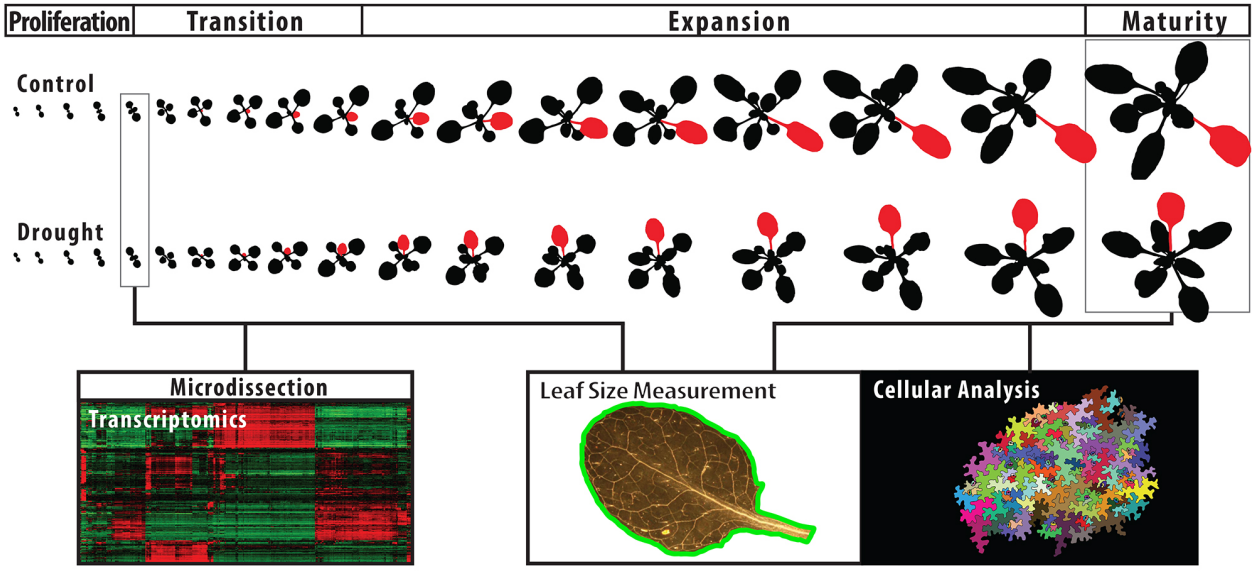
The online tool CORNET (https://cornet.psb.ugent.be/; De Bodt et al., 2010) was used to perform co-expression analysis. Four sets of microarray experiments were selected: Compendium 2 (a global collection of various microarray experiments without bias to specific conditions or tissues), Leaf (microarray experiments of leaf tissue), Development (microarray experiments from different tissues, developmental stages, and developmental mutants), and Stress (microarray experiments under various biotic and abiotic stresses). A Pearson correlation cutoff of 0.7 was used for the co-expression analysis. Only co-expression between the 354 common drought genes was allowed. The four obtained co-expression networks were merged and visualized in Cytoscape (Cline et al., 2007). Regulatory interactions were retrieved with the TF tool in CORNET. Interactions between query genes and neighbors were included. From the AGRIS database, only confirmed interactions were used. Computationally inferred regulatory interactions from microarray data were added to the network together with regulatory interactions retrieved from the literature using the EVEX text-mining tool (Van Landeghem et al., 2012). The network of regulatory interactions was merged with the co-expression network using Cytoscape.

Transcriptome (RNA-Sequencing) data from this article can be found in the ArrayExpress data libraries under accession number E-MTAB-3279.

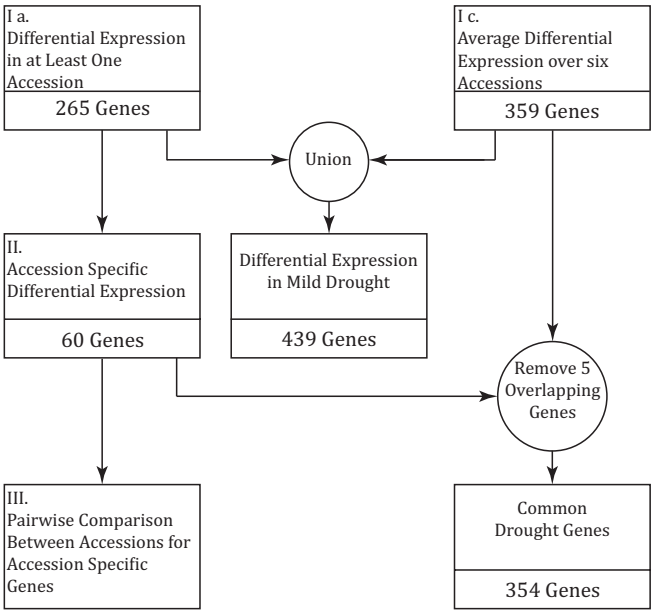
Acknowledgements

We thank the entire Systems Biology of Yield group for creating a stimulating scientific environment with many fruitful discussions. We also thank Dr. Annick Bleys for help in improving the article and Marieke Dubois for fruitful discussions.

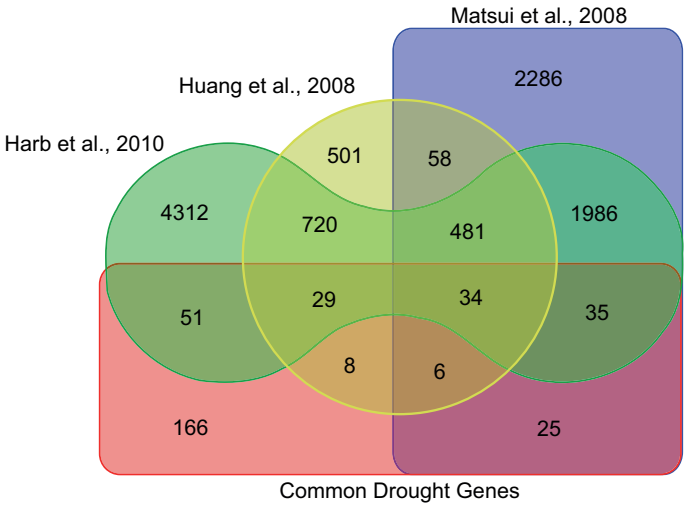
Supplementary figures



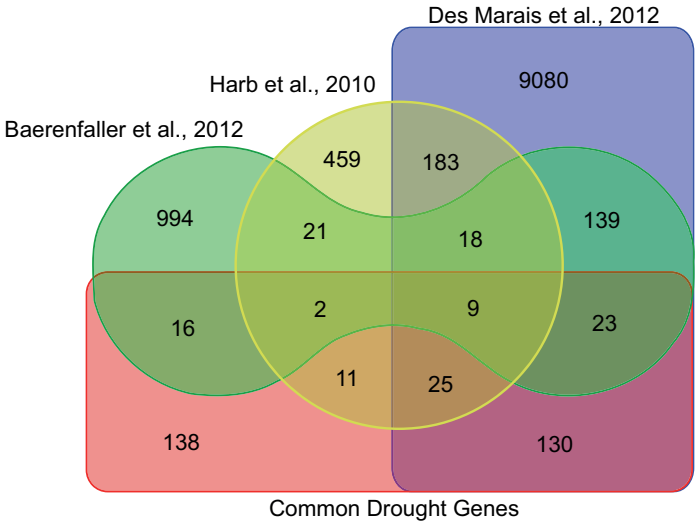
Supplementary Figure 1: Overview of the experimental set-up and measured growth parameters. At the beginning of the transition phase, the third leaf (indicated in red) was harvested for transcriptome profiling by RNA Sequencing. At the same time point and at maturity, the size of the third leaf was measured. The mature leaves were also used for a cellular analysis. By growing plants on the WIWAM, time series of the projected rosette area were obtained. All measurements were performed on plants grown in both control and mild drought conditions.



Supplementary Figure 2: Schematic overview of the three stage statistical analysis of the transcriptome data. Stage I.a (testing for differential expression of each gene in at least one accession) and stage I.c (testing for average differential expression over all six accessions for each gene) delivered, by taking the union, 439 genes to be differentially expressed under mild drought. The accessions specificity (i.e. accession x treatment interaction) of the differential expression of each gene found in stage I.a is tested in stage II. In stage III, finally, for all 60 accession specific genes the pairwise comparisons between all accessions were made. Since 5 genes were found both to be accession specific (stage II) and to be on average differentially expressed (stage I.c), these 5 genes were removed from the set of common drought genes.



Supplementary Figure 3: Venn diagram showing the overlap between the common drought genes and the selected severe drought studies (Huang et al., 2008; Matsui et al., 2008; Harb et al., 2010).



Supplementary Figure 4: Venn diagram showing the overlap between the common drought genes and the selected severe drought studies (Harb et al., 2010; Baerenfaller et al., 2012; Des Marais et al., 2012).

Supplementary data

Accession	Region	Country
An-1	Western Europe	Belgium
Blh-1	Eastern Europe	Czech Republic
Col-0	North America	U.S. of America
Cvi-0	Macaronesia	Republic of Cape Verde
Oy-0	Northern Europe	Norway
Sha	Central Asia	Tadjikistan

Supplementary Table 1. Geographic origins of the six Arabidopsis accessions.

	Comparsion	p-value	FDR corrected p-value		Comparison	p-value	FDR corrected p-value		Comparison	p-value	FDR corrected p-value
PRA Control vs. Mild Drought over Time	An-1 vs. Blh-1	0.1272	0.2057	Leaf 3 Control vs. Mild Drought at Maturity	An-1 vs. Blh-1	0.8408	0.8408	L3 Control vs. Mild Drought at Proliferation	An-1 vs. Blh-1	0.9642	0.9642
	An-1 vs. Col-0	0.8461	0.9763		An-1 vs. Col-0	0.4002	0.4617		An-1 vs. Col-0	0.1921	0.3202
	An-1 vs. Cvi-0	0.1063	0.1994		An-1 vs. Cvi-0	0.1286	0.1929		An-1 vs. Cvi-0	0.1612	0.3202
	An-1 vs. Oy-0	0.1013	0.1994		**An-1 vs. Oy-0**	**<.0001**	**0.0003**		An-1 vs. Oy-0	0.3283	0.3918
	An-1 vs. Sha	0.1371	0.2057		**An-1 vs. Sha**	**0.0018**	**0.0055**		An-1 vs. Sha	0.0119	0.0595
	Blh-1 vs. Col-0	0.1823	0.2433		Blh-1 vs. Col-0	0.3054	0.4164		Blh-1 vs. Col-0	0.2407	0.361
	Blh-1 vs. Cvi-0	**0.0024**	**0.0104**		Blh-1 vs. Cvi-0	0.0914	0.1713		Blh-1 vs. Cvi-0	0.176	0.3202
	Blh-1 vs. Oy-0	**0.0016**	**0.0104**		**Blh-1 vs. Oy-0**	**<.0001**	**0.0003**		Blh-1 vs. Oy-0	0.3395	0.3918
	Blh-1 vs. Sha	0.9818	0.9818		**Blh-1 vs. Sha**	**0.0012**	**0.0043**		Blh-1 vs. Sha	0.0169	0.0634
	Col-0 vs. Cvi-0	0.072	0.1801		Col-0 vs. Cvi-0	0.4935	0.5288		Col-0 vs. Cvi-0	0.0103	0.0595
	Col-0 vs. Oy-0	0.0666	0.1801		**Col-0 vs. Oy-0**	**0.0011**	**0.0043**		Col-0 vs. Oy-0	0.0299	0.0896
	Col-0 vs. Sha	0.1946	0.2433		**Col-0 vs. Sha**	**0.0215**	**0.0461**		**Col-0 vs. Sha**	**0.0003**	**0.0039**
	Cvi-0 vs. Oy-0	0.9407	0.9818		**Cvi-0 vs. Oy-0**	**0.0102**	**0.0254**		Cvi-0 vs. Oy-0	0.6793	0.7278
	Cvi-0 vs. Sha	**0.0028**	**0.0104**		Cvi-0 vs. Sha	0.1054	0.1757		Cvi-0 vs. Sha	0.3074	0.3918
	Oy-0 vs. Sha	**0.0019**	**0.0104**		Oy-0 vs. Sha	0.36	0.45		Oy-0 vs. Sha	0.1463	0.3202
PRA Control vs. Mild Drought at Maturity	An-1 vs. Blh-1	0.5891	0.7364	Leaf 3 Control at Maturity	An-1 vs. Blh-1	0.883	0.883	L3 Control at Proliferation	An-1 vs. Blh-1	0.4002	0.4288
	An-1 vs. Col-0	0.6564	0.7574		**An-1 vs. Col-0**	**0.0085**	**0.0127**		An-1 vs. Col-0	0.0417	0.0782
	An-1 vs. Cvi-0	0.1224	0.2623		An-1 vs. Cvi-0	0.1345	0.1552		**An-1 vs. Cvi-0**	**<.0001**	**<.0001**
	An-1 vs. Oy-0	**0.002**	**0.0099**		**An-1 vs. Oy-0**	**<.0001**	**<.0001**		An-1 vs. Oy-0	0.075	0.1073
	An-1 vs. Sha	0.7843	0.7939		**An-1 vs. Sha**	**<.0001**	**<.0001**		An-1 vs. Sha	0.8091	0.8091
	Blh-1 vs. Col-0	0.3269	0.4903		**Blh-1 vs. Col-0**	**0.0066**	**0.011**		**Blh-1 vs. Col-0**	**0.0068**	**0.0145**
	Blh-1 vs. Cvi-0	0.0415	0.1244		Blh-1 vs. Cvi-0	0.1077	0.1454		**Blh-1 vs. Cvi-0**	**<.0001**	**<.0001**
	Blh-1 vs. Oy-0	**0.0003**	**0.0048**		**Blh-1 vs. Oy-0**	**<.0001**	**<.0001**		Blh-1 vs. Oy-0	0.398	0.4288
	Blh-1 vs. Sha	0.7939	0.7939		**Blh-1 vs. Sha**	**<.0001**	**<.0001**		Blh-1 vs. Sha	0.2948	0.3686
	Col-0 vs. Cvi-0	0.2599	0.4331		Col-0 vs. Cvi-0	0.2555	0.2738		**Col-0 vs. Cvi-0**	**<.0001**	**<.0001**
	Col-0 vs. Oy-0	**0.0079**	**0.0297**		**Col-0 vs. Oy-0**	**<.0001**	**<.0001**		**Col-0 vs. Oy-0**	**0.0002**	**0.0006**
	Col-0 vs. Sha	0.4765	0.6497		**Col-0 vs. Sha**	**0.0008**	**0.0016**		Col-0 vs. Sha	0.0787	0.1073
	Cvi-0 vs. Oy-0	0.1671	0.3134		**Cvi-0 vs. Oy-0**	**<.0001**	**<.0001**		**Cvi-0 vs. Oy-0**	**<.0001**	**<.0001**
	Cvi-0 vs. Sha	0.0748	0.1869		**Cvi-0 vs. Sha**	**<.0001**	**<.0001**		**Cvi-0 vs. Sha**	**<.0001**	**<.0001**
	Oy-0 vs. Sha	**0.0009**	**0.0068**		Oy-0 vs. Sha	0.1163	0.1454		Oy-0 vs. Sha	0.0487	0.0811
PRA Control at Maturity	**An-1 vs. Blh-1**	**0.0272**	**0.0371**	Leaf 3 Mild Drought at Maturity	An-1 vs. Blh-1	0.8909	0.9546	L3 Mild Drought at Proliferation	An-1 vs. Blh-1	0.5254	0.563
	An-1 vs. Col-0	**0.0006**	**0.0013**		An-1 vs. Col-0	0.1663	0.3563		**An-1 vs. Col-0**	**0.0007**	**0.0013**
	An-1 vs. Cvi-0	0.0422	0.0528		An-1 vs. Cvi-0	0.5016	0.684		**An-1 vs. Cvi-0**	**<.0001**	**<.0001**
	An-1 vs. Oy-0	**<.0001**	**<.0001**		An-1 vs. Oy-0	0.0973	0.3563		**An-1 vs. Oy-0**	**0.0078**	**0.0117**
	An-1 vs. Sha	**0.0017**	**0.0032**		An-1 vs. Sha	0.161	0.3563		**An-1 vs. Sha**	**0.0025**	**0.0042**
	Blh-1 vs. Col-0	0.2389	0.2757		Blh-1 vs. Col-0	0.2218	0.3697		**Blh-1 vs. Col-0**	**0.0002**	**0.0004**
	Blh-1 vs. Cvi-0	**<.0001**	**0.0001**		Blh-1 vs. Cvi-0	0.4258	0.6387		**Blh-1 vs. Cvi-0**	**<.0001**	**<.0001**
	Blh-1 vs. Oy-0	**0.0002**	**0.0005**		Blh-1 vs. Oy-0	0.1355	0.3563		Blh-1 vs. Oy-0	0.0589	0.0736
	Blh-1 vs. Sha	0.3642	0.3902		Blh-1 vs. Sha	0.214	0.3697		**Blh-1 vs. Sha**	**0.027**	**0.0368**
	Col-0 vs. Cvi-0	**<.0001**	**<.0001**		Col-0 vs. Cvi-0	0.0404	0.2022		Col-0 vs. Cvi-0	0.0807	0.0931
	Col-0 vs. Oy-0	**0.0104**	**0.0156**		Col-0 vs. Oy-0	0.77	0.9382		**Col-0 vs. Oy-0**	**<.0001**	**<.0001**
	Col-0 vs. Sha	0.7909	0.7909		Col-0 vs. Sha	0.9603	0.9603		**Col-0 vs. Sha**	**<.0001**	**<.0001**
	Cvi-0 vs. Oy-0	**<.0001**	**<.0001**		Cvi-0 vs. Oy-0	0.0207	0.2022		**Cvi-0 vs. Oy-0**	**<.0001**	**<.0001**
	Cvi-0 vs. Sha	**<.0001**	**<.0001**		Cvi-0 vs. Sha	0.0403	0.2022		**Cvi-0 vs. Sha**	**<.0001**	**<.0001**
	Oy-0 vs. Sha	**0.005**	**0.0083**		Oy-0 vs. Sha	0.8131	0.9382		Oy-0 vs. Sha	0.7502	0.7502

Supplementary Table 2. Pairwise comparisons of projected rosette area and leaf 3 size measurements. Significant differences are indictated in bold ($P < 0.05$).

Due to their large size, the following tables are provided electronically via:
http://www.plantphysiol.org/content/167/3/800.long

Supplementary Table 3. List of differential expressed genes.

Supplementary Table 4. List of enriched GO categories.

Supplementary Table 5. Common drought genes previously associated with ABA.

Supplementary Table 6. Putative DELLA targets among the common drought genes.

Supplementary Table 7. Common drought genes not previously anno- tated to be drought responsive.

Supplementary Table 8. Top 20 interactors from the coexpression and transcriptional regulation network.

Supplementary Table 9. Eighty-seven genes specific for young developing tissue exposed to mild drought.

References

Achard P, Cheng H, De Grauwe L, Decat J, Schoutteten H, Moritz T, Van Der Straeten D, Peng J, Harberd NP (2006) Integration of plant responses to environmentally activated phytohormonal signals. Science 311: 91–94

Achard P, Gong F, Cheminant S, Alioua M, Hedden P, Genschik P (2008) The cold-inducible CBF1 factor-dependent signaling pathway modulates the accumulation of the growth-repressing DELLA proteins via its effect on gibberellin metabolism. Plant Cell 20: 2117–2129

Aguirrezabal L, Bouchier-Combaud S, Radziejwoski A, Dauzat M, Cookson SJ, Granier C (2006) Plasticity to soil water deficit in Arabidopsis thaliana: dissection of leaf development into underlying growth dynamic and cellular variables reveals invisible phenotypes. Plant Cell Environ 29: 2216–2227

Akhtar M, Jaiswal A, Taj G, Jaiswal JP, Qureshi MI, Singh NK (2012) DREB1/CBF transcription factors: their structure, function and role in abiotic stress tolerance in plants. J Genet 91: 385–395

Anastasiou E, Kenz S, Gerstung M, MacLean D, Timmer J, Fleck C, Lenhard M (2007) Control of plant organ size by KLUH/CYP78A5-dependent intercellular signaling. Dev Cell 13: 843–856

Anders S, Pyl PT, Huber W (2015) HTSeq: a Python framework to work with high-throughput sequencing data. Bioinformatics 31: 166–169

Anderson JP, Badruzsaufari E, Schenk PM, Manners JM, Desmond OJ, Ehlert C, Maclean DJ, Ebert PR, Kazan K (2004) Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in Arabidopsis. Plant Cell 16: 3460–3479

Andriankaja M, Dhondt S, De Bodt S, Vanhaeren H, Coppens F, De Milde L, Mühlenbock P, Skirycz A, Gonzalez N, Beemster GTS, et al (2012) Exit from proliferation during leaf development in *Arabidopsis thaliana*: a not-so-gradual process. Dev Cell 22: 64–78

Baerenfaller K, Massonnet C, Walsh S, Baginsky S, Bühlmann P, Hennig L, Hirsch-Hoffmann M, Howell KA, Kahlau S, Radziejwoski A, et al (2012) Systems-based analysis of Arabidopsis leaf growth reveals adaptation to water deficit. Mol Syst Biol 8: 606

Beaudoin N, Serizet C, Gosti F, Giraudat J (2000) Interactions between abscisic acid and ethylene signaling cascades. Plant Cell 12: 1103–1115

Beckers GJM, Jaskiewicz M, Liu Y, Underwood WR, He SY, Zhang S, Conrath U (2009) Mitogen-activated protein kinases 3 and 6 are required for full priming of stress responses in Arabidopsis thaliana. Plant Cell 21: 944–953

Bell E, Creelman RA, Mullet JE (1995) A chloroplast lipoxygenase is required for wound-induced jasmonic acid accumulation in Arabidopsis. Proc Natl Acad Sci USA 92: 8675–8679

Benschop JJ, Jackson MB, Gühl K, Vreeburg RAM, Croker SJ, Peeters AJM, Voesenek LACJ (2005) Contrasting interactions between ethylene and abscisic acid in Rumex species differing in submergence tolerance. Plant J 44: 756–768

Bergmann DC, Sack FD (2007) Stomatal development. Annu Rev Plant Biol 58: 163–181

Bouchabke O, Chang F, Simon M, Voisin R, Pelletier G, Durand-Tardif M (2008) Natural variation in Arabidopsis

thaliana as a tool for highlighting differential drought responses. PLoS ONE 3: e1705

Chardon F, Barthélémy J, Daniel-Vedele F, Masclaux-Daubresse C (2010) Natural variation of nitrate uptake and nitrogen use efficiency in Arabidopsis thaliana cultivated with limiting and ample nitrogen supply. J Exp Bot 61: 2293–2302

Chen WJ, Chang SH, Hudson ME, Kwan WK, Li J, Estes B, Knoll D, Shi L, Zhu T (2005) Contribution of transcriptional regulation to natural variations in Arabidopsis. Genome Biol 6: R32

Chen Z, Zhang H, Jablonowski D, Zhou X, Ren X, Hong X, Schaffrath R, Zhu JK, Gong Z (2006) Mutations in ABO1/ELO2, a subunit of holoElongator, increase abscisic acid sensitivity and drought tolerance in Arabidopsis thaliana. Mol Cell Biol 26: 6902–6912

Cheng WH, Chiang MH, Hwang SG, Lin PC (2009) Antagonism between abscisic acid and ethylene in Arabidopsis acts in parallel with the reciprocal regulation of their metabolism and signaling pathways. Plant Mol Biol 71: 61–80

Cheong YH, Kim KN, Pandey GK, Gupta R, Grant JJ, Luan S (2003) CBL1, a calcium sensor that differentially regulates salt, drought, and cold responses in Arabidopsis. Plant Cell 15: 1833–1845

Cheong YH, Pandey GK, Grant JJ, Batistic O, Li L, Kim BG, Lee SC, Kudla J, Luan S (2007) Two calcineurin B-like calcium sensors, interacting with protein kinase CIPK23, regulate leaf transpiration and root potassium uptake in Arabidopsis. Plant J 52: 223–239

Claeys H, De Bodt S, Inzé D (2014a) Gibberellins and DELLAs: central nodes in growth regulatory networks. Trends Plant Sci 19: 231–239 Claeys H, Inzé D (2013) The agony of choice: how plants balance growth and survival under water-limiting conditions. Plant Physiol 162: 1768– 1779

Claeys H, Skirycz A, Maleux K, Inzé D (2012) DELLA signaling mediates stress-induced cell differentiation in Arabidopsis leaves through modulation of anaphase-promoting complex/cyclosome activity. Plant Physiol 159: 739–747

Claeys H, Van Landeghem S, Dubois M, Maleux K, Inzé D (2014b) What is stress? Dose-response effects in commonly used in vitro stress assays. Plant Physiol 165: 519–527

Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, et al (2007) Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2: 2366–2382

Cominelli E, Galbiati M, Vavasseur A, Conti L, Sala T, Vuylsteke M, Leonhardt N, Dellaporta SL, Tonelli C (2005) A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. Curr Biol 15: 1196–1200

Cosgrove DJ (2005) Growth of the plant cell wall. Nat Rev Mol Cell Biol 6: 850–861

Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR (2010) Abscisic acid: emergence of a core signaling network. Annu Rev Plant Biol 61: 651–679

Danisman S, van der Wal F, Dhondt S, Waites R, de Folter S, Bimbo A, van Dijk ADJ, Muino JM, Cutri L, Dornelas MC, et al (2012) Arabidopsis class I and class II TCP transcription factors regulate jasmonic acid metabolism and leaf development antagonistically. Plant Physiol 159: 1511–1523

Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, Grotewold E (2003) AGRIS: Arabidopsis Gene Regulatory Information Server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics 4: 25

De Bodt S, Carvajal D, Hollunder J, Van den Cruyce J, Movahedi S, Inzé D (2010) CORNET: a user-friendly tool for data mining and integration. Plant Physiol 152: 1167–1179

De Bodt S, Hollunder J, Nelissen H, Meulemeester N, Inzé D (2012) CORNET 2.0: integrating plant coexpression, protein-protein interactions, regulatory interactions, gene associations and functional annotations. New Phytol 195: 707–720

Delker C, Pöschl Y, Raschke A, Ullrich K, Ettingshausen S, Hauptmann V, Grosse I, Quint M (2010) Natural variation of transcriptional auxin response networks in Arabidopsis thaliana. Plant Cell 22: 2184–2200

De Paepe A, Vuylsteke M, Van Hummelen P, Zabeau M, Van Der Straeten D (2004) Transcriptional profiling by cDNA-AFLP and microarray analysis reveals novel insights into the early response to ethylene in Arabidopsis. Plant J 39: 537–559

Des Marais DL, McKay JK, Richards JH, Sen S, Wayne T, Juenger TE (2012) Physiological genomics of response to soil drying in diverse Arabidopsis accessions. Plant Cell 24: 893–914

Donnelly PM, Bonetta D, Tsukaya H, Dengler RE, Dengler NG (1999) Cell cycling and cell enlargement in developing leaves of Arabidopsis. Dev Biol 215: 407–419

Dubois M, Skirycz A, Claeys H, Maleux K, Dhondt S, De Bodt S, Vanden Bossche R, De Milde L, Yoshizumi T, Matsui M, et al (2013) Ethylene Response Factor6 acts as a central regulator of leaf growth under water-limiting conditions in Arabidopsis. Plant Physiol 162: 319–332

Eriksson S, Stransfeld L, Adamski NM, Breuninger H, Lenhard M (2010) KLUH/CYP78A5-dependent growth signaling coordinates floral organ growth in Arabidopsis. Curr Biol 20: 527–532

Finkelstein R (2013) Abscisic acid synthesis and response. The Arabidopsis Book 11: e0166, doi/10.1199/tab.0166

Geisler M, Nadeau J, Sack FD (2000) Oriented asymmetric divisions that generate the stomatal spacing pattern in Arabidopsis are disrupted by the too many mouths mutation. Plant Cell 12: 2075–2086

Ghassemian M, Nambara E, Cutler S, Kawaide H, Kamiya Y, McCourt P (2000) Regulation of abscisic acid signaling by the ethylene response pathway in Arabidopsis. Plant Cell 12: 1117–1126

Golldack D, Li C, Mohan H, Probst N (2014) Tolerance to drought and salt stress in plants: unraveling the signaling networks. Front Plant Sci 5: 151

Gómez-Cadenas A, Zentella R, Walker-Simmons MK, Ho THD (2001) Gibberellin/abscisic acid antagonism in barley aleurone cells: site of action of the protein kinase PKABA1 in relation to gibberellin signaling molecules. Plant Cell 13: 667–679

Gonzalez N, Vanhaeren H, Inzé D (2012) Leaf size control: complex coordination of cell division and expansion. Trends Plant Sci 17: 332–340

Harb A, Krishnan A, Ambavaram MMR, Pereira A (2010) Molecular and physiological analysis of drought stress in Arabidopsis reveals early responses leading to acclimation in plant growth. Plant Physiol 154: 1254–1271

Hoffmann MH (2002) Biogeography of Arabidopsis thaliana (L.) Heynh. (Brassicaceae). J Biogeogr 29: 125–134 Huang D, Wu W, Abrams SR, Cutler AJ (2008) The relationship of drought-related gene expression in Arabidopsis thaliana to hormonal and environmental factors. J Exp Bot 59: 2991–3007

Iuchi S, Kobayashi M, Taji T, Naramoto M, Seki M, Kato T, Tabata S, Kakubari Y, Yamaguchi-Shinozaki K, Shinozaki K (2001) Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in Arabidopsis. Plant J 27: 325–333

Jiang H, Doerge RW (2006) A two-step multiple comparison procedure for a large number of tests and multiple treatments. Stat Appl Genet Mol Biol 5: e28

Katori T, Ikeda A, Iuchi S, Kobayashi M, Shinozaki K, Maehashi K, Sakata Y, Tanaka S, Taji T (2010) Dissecting the genetic control of natural variation in salt tolerance of Arabidopsis thaliana accessions. J Exp Bot 61: 1125–1138

Kim SH, Woo DH, Kim JM, Lee SY, Chung WS, Moon YH (2011) Arabidopsis MKK4 mediates osmotic-stress response via its regulation of MPK3 activity. Biochem Biophys Res Commun 412: 150–154

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in Arabidopsis thaliana. Annu Rev Plant Biol 55: 141–172

Lawlor DW (2013) Genetic engineering to improve plant performance under drought: physiological evaluation of achievements, limitations, and possibilities. J Exp Bot 64: 83–108

Lechner L, Pereyra-Irujo GA, Granier C, Aguirrezábal LAN (2008) Rewatering plants after a long water-deficit treatment reveals that leaf epidermal cells retain their ability to expand after the leaf has apparently reached its final size. Ann Bot (Lond) 101: 1007–1015

Lee SJ, Kang JY, Park HJ, Kim MD, Bae MS, Choi HI, Kim SY (2010) DREB2C interacts with ABF2, a bZIP protein regulating abscisic acidresponsive gene expression, and its overexpression affects abscisic acid sensitivity. Plant Physiol 153: 716–727

Li C, Potuschak T, Colón-Carmona A, Gutiérrez RA, Doerner P (2005) Arabidopsis TCP20 links regulation of growth and cell division control pathways. Proc Natl Acad Sci USA 102: 12978–12983

Li G, Meng X, Wang R, Mao G, Han L, Liu Y, Zhang S (2012) Dual-level regulation of ACC synthase activity by MPK3/MPK6 cascade and its downstream WRKY transcription factor during ethylene induction in

Arabidopsis. PLoS Genet 8: e1002767

Llorente F, López-Cobollo RM, Catalá R, Martínez-Zapater JM, Salinas J (2002) A novel cold-inducible gene from Arabidopsis, RCI3, encodes a peroxidase that constitutes a component for stress tolerance. Plant J 32: 13–24 Lü P, Zhang C, Liu J, Liu X, Jiang G, Jiang X, Khan MA, Wang L, Hong B,

Gao J (2014) RhHB1 mediates the antagonism of gibberellins to ABA and ethylene during rose (Rosa hybrida) petal senescence. Plant J 78: 578– 590

MacMillan CP, Mansfield SD, Stachurski ZH, Evans R, Southerton SG (2010) Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in Arabidopsis and Eucalyptus. Plant J 62: 689–703

Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. Bioinformatics 21: 3448–3449

Magome H, Yamaguchi S, Hanada A, Kamiya Y, Oda K (2008) The DDF1 transcriptional activator upregulates expression of a gibberellin-deactivating gene, GA2ox7, under high-salinity stress in Arabidopsis. Plant J 56: 613–626

Matsui A, Ishida J, Morosawa T, Mochizuki Y, Kaminuma E, Endo TA, Okamoto M, Nambara E, Nakajima M, Kawashima M, et al (2008) Arabidopsis transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. Plant Cell Physiol 49: 1135–1149

McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 40: 4288–4297

McDowell N, Pockman WT, Allen CD, Breshears DD, Cobb N, Kolb T, Plaut J, Sperry J, West A, Williams DG, et al (2008) Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? New Phytol 178: 719–739

Meng X, Xu J, He Y, Yang KY, Mordorski B, Liu Y, Zhang S (2013) Phosphorylation of an ERF transcription factor by Arabidopsis MPK3/ MPK6 regulates plant defense gene induction and fungal resistance. Plant Cell 25: 1126–1142

Miyakawa T, Fujita Y, Yamaguchi-Shinozaki K, Tanokura M (2013) Structure and function of abscisic acid receptors. Trends Plant Sci 18: 259–266

Moore JP, Vicré-Gibouin M, Farrant JM, Driouich A (2008) Adaptations of higher plant cell walls to water loss: drought vs desiccation. Physiol Plant 134: 237–245

Nakagami H, Pitzschke A, Hirt H (2005) Emerging MAP kinase pathways in plant stress signalling. Trends Plant Sci 10: 339–346

Navarro L, Bari R, Achard P, Lisón P, Nemri A, Harberd NP, Jones JDG (2008) DELLAs control plant immune responses by modulating the balance of jasmonic acid and salicylic acid signaling. Curr Biol 18: 650– 655

Nelson DE, Repetti PP, Adams TR, Creelman RA, Wu J, Warner DC, Anstrom DC, Bensen RJ, Castiglioni PP, Donnarummo MG, et al (2007) Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc Natl Acad Sci USA 104: 16450–16455

Nemhauser JL, Hong F, Chory J (2006) Different plant hormones regulate similar processes through largely nonoverlapping transcriptional responses. Cell 126: 467–475

Osnato M, Castillejo C, Matías-Hernández L, Pelaz S (2012) TEMPRANILLO genes link photoperiod and gibberellin pathways to control flowering in Arabidopsis. Nat Commun 3: 808

Palaniswamy SK, James S, Sun H, Lamb RS, Davuluri RV, Grotewold E (2006) AGRIS and AtRegNet: a platform to link cis-regulatory elements and transcription factors into regulatory networks. Plant Physiol 140: 818–829

Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC, Weigel D (2003) Control of leaf morphogenesis by microRNAs. Nature 425: 257–263

Park JE, Park JY, Kim YS, Staswick PE, Jeon J, Yun J, Kim SY, Kim J, Lee YH, Park CM (2007) GH3-mediated auxin homeostasis links growth regulation with stress adaptation response in Arabidopsis. J Biol Chem 282: 10036–10046

Pauwels L, Morreel K, De Witte E, Lammertyn F, Van Montagu M, Boerjan W, Inzé D, Goossens A (2008) Mapping methyl jasmonate-mediated transcriptional reprogramming of metabolism and cell cycle progression in cultured Arabidopsis cells. Proc Natl Acad Sci USA 105: 1380–1385

Pedley KF, Martin GB (2005) Role of mitogen-activated protein kinases in plant immunity. Curr Opin Plant Biol 8: 541–547

Pereyra-Irujo GA, Velázquez L, Lechner L, Aguirrezábal LAN (2008) Genetic variability for leaf growth rate and duration under water deficit in sunflower: analysis of responses at cell, organ, and plant level. J Exp Bot 59: 2221–2232

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140

Sarvepalli K, Nath U (2011) Hyper-activation of the TCP4 transcription factor in Arabidopsis thaliana accelerates multiple aspects of plant maturation. Plant J 67: 595–607

Schaalje GB, McBride JB, Fellingham GW (2001) Approximations to distributions of test statistics in complex mixed linear models using SAS Proc MIXED. In Proceedings of SAS User Group International 26th Annual Conference, Long Beach, CA, April 22–25 2001, paper 262-26. SAS Institute Inc., Cary, NC

Skirycz A, Claeys H, De Bodt S, Oikawa A, Shinoda S, Andriankaja M, Maleux K, Eloy NB, Coppens F, Yoo SD, et al (2011a) Pause-and-stop: the effects of osmotic stress on cell proliferation during early leaf development in Arabidopsis and a role for ethylene signaling in cell cycle arrest. Plant Cell 23: 1876–1888

Skirycz A, De Bodt S, Obata T, De Clercq I, Claeys H, De Rycke R, Andriankaja M, Van Aken O, Van Breusegem F, Fernie AR, et al (2010) Developmental stage specificity and the role of mitochondrial metabolism in the response of Arabidopsis leaves to prolonged mild osmotic stress. Plant Physiol 152: 226–244

Skirycz A, Vandenbroucke K, Clauw P, Maleux K, De Meyer B, Dhondt S, Pucci A, Gonzalez N, Hoeberichts F, Tognetti VB, et al (2011b) Survival and growth of Arabidopsis plants given limited water are not equal. Nat Biotechnol 29: 212–214

Strizhov N, Abrahám E, Okrész L, Blickling S, Zilberstein A, Schell J, Koncz C, Szabados L (1997) Differential expression of two *P5CS* genes controlling proline accumulation during salt-stress requires ABA and is regulated by ABA1, ABI1 and AXR2 in Arabidopsis. Plant J 12: 557–569

Swiątek A, Azmi A, Stals H, Inzé D, Van Onckelen H (2004) Jasmonic acid prevents the accumulation of cyclin B1;1 and CDK-B in synchronized tobacco BY-2 cells. FEBS Lett 572: 118–122

Swiątek A, Lenjou M, Van Bockstaele D, Inzé D, Van Onckelen H (2002) Differential effect of jasmonic acid and abscisic acid on cell cycle progression in tobacco BY-2 cells. Plant Physiol 128: 201–211

Szabados L, Savouré A (2010) Proline: a multifunctional amino acid. Trends Plant Sci 15: 89–97

Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K (2002) Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. Plant J 29: 417–426

Tan BC, Joseph LM, Deng WT, Liu L, Li QB, Cline K, McCarty DR (2003) Molecular characterization of the Arabidopsis 9-cis epoxycarotenoid dioxygenase gene family. Plant J 35: 44–56

Tanaka Y, Sano T, Tamaoki M, Nakajima N, Kondo N, Hasezawa S (2005) Ethylene inhibits abscisic acid-induced stomatal closure in Arabidopsis. Plant Physiol 138: 2337–2343

Tardieu F (2012) Any trait or trait-related allele can confer drought tolerance: just design the right drought scenario. J Exp Bot 63: 25–31

Tardieu F, Parent B, Simonneau T (2010) Control of leaf growth by abscisic acid: hydraulic or non-hydraulic processes? Plant Cell Environ 33: 636–647

Tisné S, Reymond M, Vile D, Fabre J, Dauzat M, Koornneef M, Granier C (2008) Combined genetic and modeling approaches reveal that epidermal cell area and number in leaves are controlled by leaf and plant developmental processes in Arabidopsis. Plant Physiol 148: 1117–1127

Tran LSP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2004) Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. Plant Cell 16: 2481–2498

Trontin C, Kiani S, Corwin JA, Hématy K, Yansouni J, Kliebenstein DJ, Loudet O (2014) A pair of receptor-like kinases is responsible for natural variation in shoot growth response to mannitol treatment in *Arabidopsis thaliana*. Plant J 78: 121–133

Umezawa T, Nakashima K, Miyakawa T, Kuromori T, Tanokura M, Shinozaki K, Yamaguchi-Shinozaki K (2010) Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. Plant Cell Physiol 51: 1821–1839

Umezawa T, Okamoto M, Kushiro T, Nambara E, Oono Y, Seki M, Kobayashi M, Koshiba T, Kamiya Y, Shinozaki K (2006) CYP707A3, a major ABA 89-hydroxylase involved in dehydration and rehydration response in *Arabidopsis thaliana*. Plant J 46: 171–182

Umezawa T, Yoshida R, Maruyama K, Yamaguchi-Shinozaki K, Shinozaki K (2004) SRK2C, a SNF1-related protein kinase 2, improves drought tolerance by controlling stress-responsive gene expression in Arabidopsis thaliana. Proc Natl Acad Sci USA 101: 17306–17311

Van Landeghem S, De Bodt S, Drebert ZJ, Inzé D, Van de Peer Y (2013) The potential of text mining in data integration and network biology for plant research: a case study on Arabidopsis. Plant Cell 25: 794–807

Van Landeghem S, Hakala K, Rönnqvist S, Salakoski T, Van de Peer Y, Ginter F (2012) Exploring biomolecular literature with EVEX: connecting genes through events, homology, and indirect associations. Adv Bioinformatics 2012: 582765

Verelst W, Skirycz A, Inzé D (2010) Abscisic acid, ethylene and gibberellic acid act at different developmental stages to instruct the adaptation of young leaves to stress. Plant Signal Behav 5: 473–475

Verslues PE, Agarwal M, Katiyar-Agarwal S, Zhu J, Zhu JK (2006) Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. Plant J 45: 523– 539

Verslues PE, Sharma S (2010) Proline metabolism and its implications for plant-environment interaction. The Arabidopsis Book 8: e0140, doi/ 10.1199/tab.0140

Wang JW, Schwab R, Czech B, Mica E, Weigel D (2008) Dual effects of miR156-targeted SPL genes and CYP78A5/KLUH on plastochron length and organ size in Arabidopsis thaliana. Plant Cell 20: 1231–1243

Wang P, Xue L, Batelli G, Lee S, Hou YJ, Van Oosten MJ, Zhang H, Tao WA, Zhu JK (2013) Quantitative phosphoproteomics identifies SnRK2 protein kinase substrates and reveals the effectors of abscisic acid action. Proc Natl Acad Sci USA 110: 11205–11210

Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26: 873–881

Wu Y, Cosgrove DJ (2000) Adaptation of roots to low water potentials by changes in cell wall extensibility and cell wall proteins. J Exp Bot 51: 1543–1553

Yamaguchi-Shinozaki K, Shinozaki K (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu Rev Plant Biol 57: 781–803

Yang S, Vanderbeld B, Wan J, Huang Y (2010) Narrowing down the targets: towards successful genetic engineering of drought-tolerant crops. Mol Plant 3: 469–490

Ye N, Zhang J (2012) Antagonism between abscisic acid and gibberellins is partially mediated by ascorbic acid during seed germination in rice. Plant Signal Behav 7: 563–565

Yilmaz A, Mejia-Guerra MK, Kurz K, Liang X, Welch L, Grotewold E (2011) AGRIS: the Arabidopsis Gene Regulatory Information Server, an update. Nucleic Acids Res 39: D1118–D1122

Yoshida T, Fujita Y, Maruyama K, Mogami J, Todaka D, Shinozaki K, Yamaguchi-Shinozaki K (2014) Four Arabidopsis AREB/ABF transcription factors function predominantly in gene expression downstream of SnRK2 kinases in abscisic acid signalling in response to osmotic stress. Plant Cell Environ 38: 35–49

Yoshida T, Fujita Y, Sayama H, Kidokoro S, Maruyama K, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2010) AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. Plant J 61: 672–685

Zhang Y, Turner JG (2008) Wound-induced endogenous jasmonates stunt plant growth by inhibiting mitosis. PLoS ONE 3: e3699

Zhang YY, Li Y, Gao T, Zhu H, Wang DJ, Zhang HW, Ning YS, Liu LJ, Wu YR, Chu CC, et al (2008) Arabidopsis SDIR1 enhances drought tolerance in crop plants. Biosci Biotechnol Biochem 72: 2251–2254

Zhu SY, Yu XC, Wang XJ, Zhao R, Li Y, Fan RC, Shang Y, Du SY, Wang XF, Wu FQ, et al (2007) Two calcium-dependent protein kinases, CPK4 and CPK11, regulate abscisic acid signal transduction in Arabidopsis. Plant Cell 19: 3019–3036

3. Association Mapping of Leaf Responses to Mild Drought Stress in *Arabidopsis thaliana*

Pieter Clauw[1,2], Frederik Coppens[1,2], Arthur Korte[3], Dorota Herman[1,2], Bram Slabbinck[1,2], Stijn Dhondt[1,2], Twiggy Van Daele[1,2], Liesbeth De Milde[1,2], Mattias Vermeersch[1,2], Katrien Maleux[1,2], Steven Maere[1,2], Nathalie Gonzalez[1,2], Dirk Inzé[1,2]

[1]Department of Plant Systems Biology, VIB, Ghent, Belgium
[2]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium
[3]Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria

AUTHOR CONTRIBUTIONS

P.C. was the main author of this chapter, performed most of the practical work and was involved in most of the data analyses (except for the stress predictors). F.C. performed the transcriptome data analysis and was involved in the GWAS and eGWAS analysis. A.K. performed the GWAS and eGWAS analyses. D.H. conducted data normalization. B.S. performed the modeling for the stress predictors. S.D. was involved in the image analysis. T.V.D., L.D.M., M.V. K.M. and P.C. were involved in the practical work. F.C., N.G. and D.I. supervised the project.

Introduction

In the previous chapter we described an experimental set-up that was effective for studying the phenotype and transcriptome responses of young developing leaves to mild drought stress. A common transcriptional response was detected between the six studied accessions, which clearly showed the involvement of ABA signaling and cell wall modifications in the mild drought response of young developing leaves. Furthermore, a number of genes were found that showed an accession-dependent expression response to the mild drought stress. These accession-specific transcription responses may explain the different phenotypic responses that were observed. However, six accessions are too few to associate the phenotype and the transcriptome with the genotype, in order to retrieve the genetic mechanisms that are involved in the natural variation in mild drought stress tolerance/sensitivity.

The phenotypes under investigation are related to growth, which can be considered as a highly complex, if not the most complex trait in plants. Due to their sessile life, plants are highly influenced by the environment and hence constantly need to adapt their growth to the fluctuating environment to keep the balance between growth and survival (Claeys and Inzé, 2013). The environment greatly affects plant size and growth through a plethora of regulatory mechanisms that work either directly or indirectly on processes that determine plant growth. In order to retrieve the genes that underlie such complex traits, a genome-wide screening is required.

Genome wide association mapping (GWAS) is a method that can analyze the correlation between the genetic differences in a collection of natural accessions and their different phenotypes. The marker SNPs that correlate significantly with the phenotype may then be linked to genes with a determining role in the observed phenotypic variation. Genome wide association studies (GWAS) have proven to be successful in detecting genes involved in wide array of traits, ranging from simple traits such as biotic stress resistance (Huard-Chauveau et al., 2013) to more complex features such as root system architecture (Rosas et al., 2013). Apart from the standard GWAS that analyzes one single trait, a multi-trait approach was proposed (Korte et al., 2012) for correlated traits, for example measurements of the same trait in contrasting environments. This method allows for detecting loci that are involved in the differential response of a trait upon a specific treatment.

In order for genetic differences to result in variable phenotypes, they may cause differences in gene functionality, but also in gene expression. Variability in gene expression can be substantial and potentially has an important function in causing phenotypic variation. For example variability in gene expression of the auxin response genes is more important than sequence polymorphisms for the natural variation in the auxin response network (Delker et al., 2010). Similarly, upon salicylic acid and mild drought, substantial variation is present in genome-wide gene expression among natural accessions (van Leeuwen et al., 2007; Marais et al., 2012). By eQTL mapping the genetic loci that associate with gene expression variation in recombinant inbred lines can be determined. Regulators of transcription can either be in physical connection with the gene (*cis*) or can be separate from the gene they regulate (*trans*). From previous studies it is suggested that *trans* regulators play a significant role in response to distinct environments, whereas *cis* regulators show

a more robust response over different conditions (Dimas et al., 2009; Drost et al., 2010; Barriere et al., 2011; Grundberg et al., 2011).

Here we present different methods to analyze the natural variation of the mild drought response of leaf growth. A collection of 98 accessions was screened for its variation in growth related phenotypes (rosette area, leaf area, cellular parameters) in response to mild drought. The variability between the accessions allowed for conducting a multi-trait GWAS to detect genes that are involved in the differential response upon mild drought. Furthermore, the transcriptome response of young developing leaves in 89 accessions was characterized. A modeling approach allowed for the identification of genes for which the expression difference upon mild drought is predictive for the stress treatment. Hence these so-called stress predictors may hold important functions in the mild drought response of young developing leaves. In the interest to get more insight in the transcriptional regulation of the gene expression response to mild drought stress, an expression GWAS (eGWAS) analysis was conducted. The eGWAS analysis showed a remarkable overrepresentation of *cis* regulatory elements in the regulation of gene expression, independent of treatment, in comparison to the regulation of differential expression upon mild drought stress. Eventually, a limited number of transcription regulating genes were detected that are potential important players in the mild drought response of young developing leaves.

Results

Phenotyping

98 Accessions from around the globe

98 accessions were selected to obtain sufficient genetic variation to study the diversity of mild drought stress responses in Arabidiopsis (Supplementary Table 1). The collection comprises accessions originating from diverse geographic regions ranging from the Cape Verdi Islands in the south to the upper north in Scandinavia and spanning the entire Eurasian continent including Japan and some North-American regions. The geographic distances between the origins of the accessions can be seen as a proxy for the genetic distance (Cao et al., 2011). For all of the different phenotypes that were measured, the genotypic differences explained a significant part of the observed variation, since the fixed factors 'accession' and the interaction 'accession × treatment' were highly significant ($P < 0.0001$) in the mixed model analysis. However, since the microenvironment of the sites from which accessions originate have not always been described in detail, it is not straightforward to link experimental observations to possible causes from the original habitat (Trontin et al., 2011). Therefore, this work will reflect little on the geographic origins of the accessions. Instead, we asked the question whether the different accessions behave differently when grown under standard and mild-drought conditions, and if so, which genetic adaptations and mechanisms are determining these differences.

Rosette and leaf size

In order to describe the growth responses of the leaves upon mild drought stress, the rosette and the third emerging leaf were measured. The quantification of the leaf growth was done non-destructively. By imaging the relatively flat Arabidopsis rosette a 2D representation of the rosette

is made. The growth related parameters of the third leaf were analyzed destructively by harvesting the plants at maturity and at the last day of full proliferation. Mild drought stress was imposed early during leaf development, as described by Clauw et al. (2015). In this protocol, seeds are first germinated under control conditions and subsequently transferred to respectively control and mild drought stress conditions at 4 DAS (Days After Stratification), the moment when the third leaf starts emerging from the shoot apical meristem. At maturity, under control conditions, both rosette and leaf 3 areas showed an almost four-fold difference between the smallest and the largest of the 98 accessions (Figure 1A, B). Spr1-6 was the smallest accession for both traits (rosette area: 283 mm^2, SE = 55 mm^2 ; leaf 3 area: 48 mm^2, SE = 8 mm^2). The accession with the largest rosette area was Lp2-2 (1057 mm^2, SE = 55 mm^2). Wa-1, the only accession of the collection that is known to be a tetraploid (Schmuths et al., 2004), had the largest leaf 3 area at maturity (185 mm^2, SE = 8 mm^2). Under mild drought conditions, similar fold changes between the smallest and largest accession are found, with rosette areas ranging from 115 mm^2 (Mz-0; SE = 56 mm^2) up to 448 mm^2 (Bor-4; SE = 60 mm^2) and leaf 3 areas from 23 mm^2 (Spr1-6; SE = 10 mm^2) up to 83 mm^2 (Kondara; SE = 10 mm^2). We also observed that rosette and leaf 3 areas at maturity are positively correlated under control conditions (r: 0.63; $P < 0.001$; Supplementary Figure 1A) and slightly better correlated under stress conditions (r: 0.76; $P < 0.001$; Supplementary Figure 1B).

Phenotype	Control	Mild drought stress	Percent reduction
rosette area (mm^2)	677 ± 7.3	257 ± 7.4	62%
leaf 3 maturity (mm^2)	121 ± 2.8	51 ± 2.9	57%
leaf 3 proliferation (mm^2)	0.027 ± 0.1	0.023 ± 0.153	18%
pavement cell area (μm^2)	3565 ± 17	2000 ± 17	42%
pavement cell number	33789 ± 64	22773 ± 64	30%
stomatal index (%)	45 ± 1.6	41 ± 1.6	6%

Table 1: Overview of the phenotyping. Averages of the phenotypes and their percent reductions under mild drought stress compared to control over the 98 accessions. Values represent averages ± SE except for leaf 3 at proliferation where the median was calculated because of the non-normal distribution of the values.

The extent to which leaf and rosette areas are reduced by the mild drought stress is highly variable between the different accessions (Figure 1A, B). On average, rosette areas are 62% (Table 1) smaller than in control conditions but reductions range from 43% (Bor-4) to 77% (N13). The average reduction of the third leaf area is in the same order of magnitude (57%; Table 1), with reductions ranging from 27% (Del-10) to 72% (ICE-61). The correlation of the relative reduction under mild drought stress with the respective sizes in control conditions is non-significant for rosette area (r = 0.10; $P = 0.33$; Supplementary Figure 1C) and very low for leaf 3 area (r = 0.39; $P < 0.0001$; Supplementary Figure 1D), showing that large plants are not necessarily more sensitive to mild drought stress. This is exemplified by Bor-4, having the least reduced rosette area but being among the 25% largest accessions. N13 on the other hand is among the 20% smallest accessions and shows the largest reduction in rosette area in response to mild drought stress. In general, large plants under control conditions also remain large under mild drought stress. This is shown by the significant positive correlation between areas in control and in mild drought stress for rosette area (r = 0.73; $P < 0.001$; Supplementary Figure 1E) and leaf 3 area at maturity (r = 0.59; $P < 0.001$; Supplementary Figure 1F).

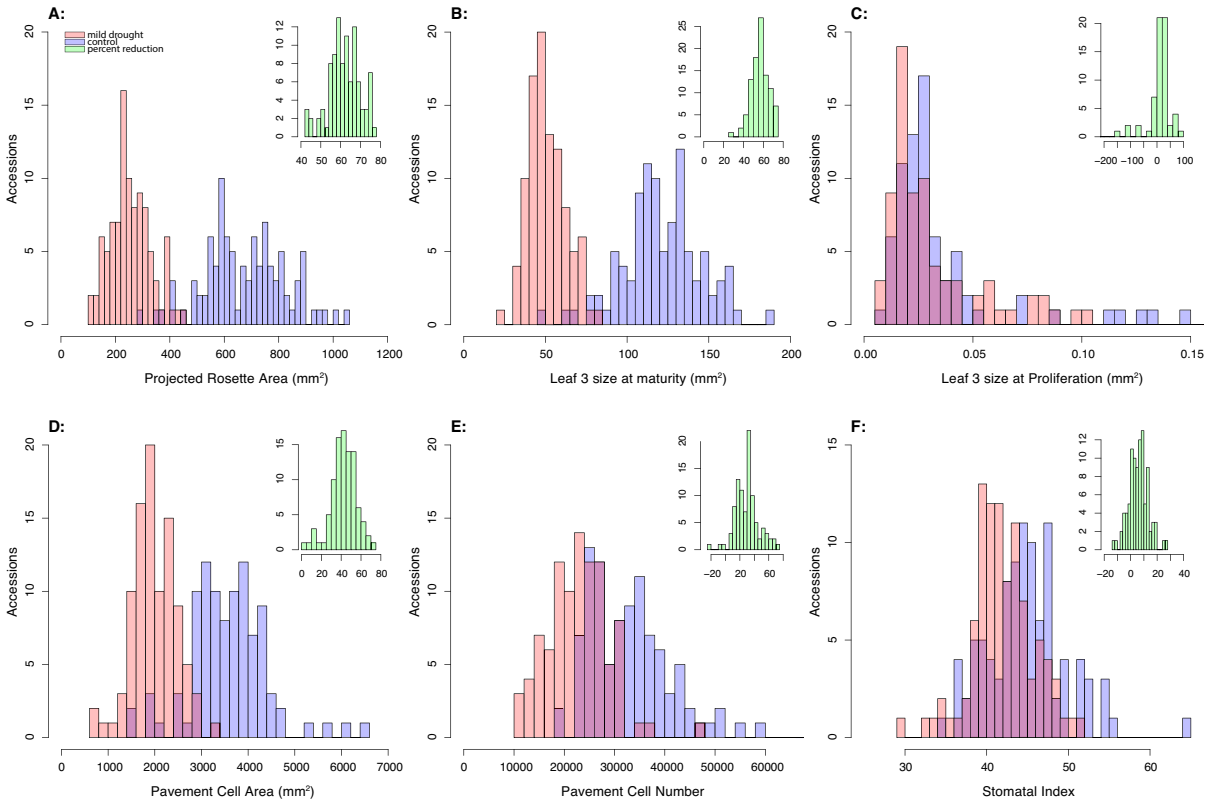


Figure 1: Distribution of leaf growth-related phenotypes in 98 accessions. Histograms of the distributions of the different phenotypes measured in 98 accessions under control (blue) and mild drought conditions (red). Distributions of percent reductions are given in the insets colored in green. A: Projected rosette area (PRA). B: leaf 3 size at maturity. C: Leaf 3 size at proliferation. D: Pavement cell area. E: Pavement cell number. F: Stomatal index. The Ler-1 accession was not included in the histogram for pavement cell number due to its extreme number of pavement cells (117165 pavement cells).

In order to have an idea on the behavior of the accessions that can cope best or worst with the mild drought stress, the most tolerant and most sensitive accessions were determined. Therefore, the ten most and the ten least reduced accessions for at one hand leaf 3 area at maturity and, on the other hand for rosette area were selected. Among these accessions, the three accessions that were among the ten least reduced for both phenotypes were NFA-10, Pna-10 and ICE-163. These three accessions are further referred to as the tolerant accessions. The three accessions that were among the ten most sensitive accessions for both phenotypes were ICE-61, Can-0 and Rubezhnoe-1, further referred to as sensitive accessions (Figure 2).

In conclusion, the automated watering and imaging systems used in this study allowed for the implementation of mild drought stress before leaf 3 emerges. In these conditions leaf growth is clearly impaired for all accessions and the response differs strongly between accessions. We also observed that tolerance and sensitivity to mild drought are not related to the plant's sizes in control conditions.

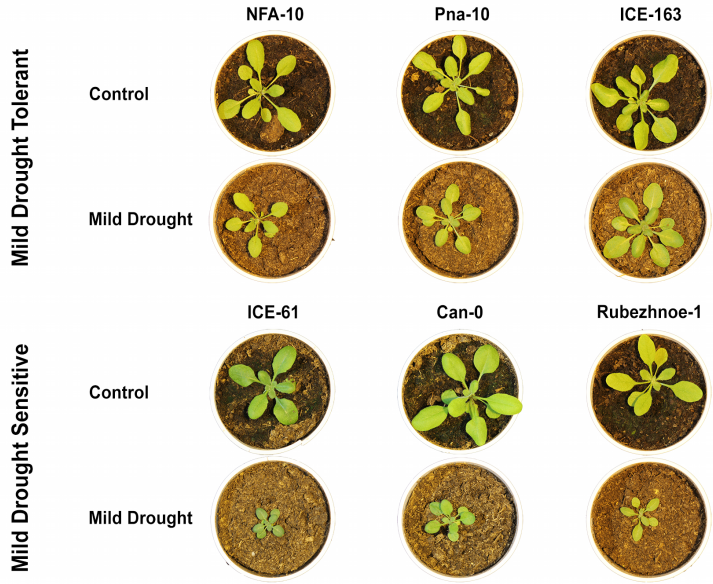


Figure 2: Rosette of mild drought tolerant and sensitive accessions. Pictures from the WIWAM platform showing the rosettes of the tolerant (ICE-163, NFA-10 and Pna-10) and sensitive accessions (Can-0, ICE-61 and Rubezhnoe-1), both in control and mild drought conditions at 22 DAS.

Cellular parameters defining leaf area

To know whether cell proliferation and/or cell expansion, the two main drivers for leaf growth, are affected by mild drought, pavement cell number and area were measured in the third leaf at maturity by making cellular drawings of the epidermis of the third leaf at maturity. These drawings were analyzed using an in-house developed algorithm (Andriankaja et al., 2012) to obtain quantitative estimates of cell number, cell areas and stomatal index.

Under control conditions, the average pavement cell area varied from 1463 μm^2 (SE = 310 μm^2) in Ler-1 to 6455 μm^2 (SE = 305 μm^2) in Sq-1, representing a more than four fold difference (Figure 1D). Ler-1 leaves, with the smallest pavement cell size, had the largest number of pavement cells (117165; SE = 4174). The accession with the second most pavement cells is Bl-1 (58531; SE = 4139), which is about twice fewer pavement cells than Ler-1. Pna-10 contained the smallest number of pavement cells (19063; SE = 4098), six times less than Ler-1 and 3 times less than Bl-1 (Figure 1E). Under mild drought stress, Ler-1 and Sq-1 are also the accessions with, respectively, the smallest (696 μm^2; SE = 363 μm^2) and largest (3204 μm^2; SE = 305 μm^2) pavement cells. The pavement cell number also differs five fold in mild drought conditions between the accessions with the most (Bl-1; 47063; SE = 4139) and the least (Yo-0; 10252; SE = 4139) pavement cells.

For both pavement cell area and pavement cell number, there was a correlation between the observations in mild drought and control conditions (resp. r = 0.55; r = 0.59; $P < 0.0001$; Supplementary Figure 1 G and H). On average, the accessions showed a 42% reduction in pavement cell area when subjected to mild drought stress, whereas the pavement cell number is 30% reduced (Table 1). The reduction in pavement cell area ranged from only 5% in ICE-138 up to 71% in Tamm-27 (Figure 1D). For pavement cell number, the largest reduction was observed for Ler-1 leaves that contain 74% less cells under stress conditions. Interestingly, three accessions, Pna-10, Br-0 and Ga-0 showed an increase in pavement cell number of respectively

20%, 9% and only 4% under mild drought stress. For two of these accessions, the reduction in pavement cell area is however much larger than the average reduction (59% for Br-0 and 63% for Ga-0) whereas in Pna-10, the reduction is of 36%. Although these three accessions produce more cells under stress, final leaf size is reduced by 65% in Br-0, 61% in Ga-0 and to a lesser extent, 28% reduction, in Pna-10. In order to find out whether reductions in pavement cell area and pavement cell number are linked either in positive (both cellular parameters contribute to the reduction in leaf area) or negative way (compensation of one cellular parameter by the other), the correlation between both parameters was calculated, but was not significant (r = -0.13; $P = 0.19$).

To know whether cell size or rather cell area is determining final leaf area in the 98 accessions, the relationship between the cellular parameters and the leaf size was further examined. Pavement cell area shows a very weak correlation with leaf 3 area (r = 0.41; $P < 0.0001$; Supplementary Figure 1I) whereas pavement cell number shows a slightly stronger correlation (r = 0.54; $P < 0.0001$; Supplementary Figure 1J). The relationship between pavement cell area, cell number and leaf 3 area under control conditions is visualised in Figure 3.

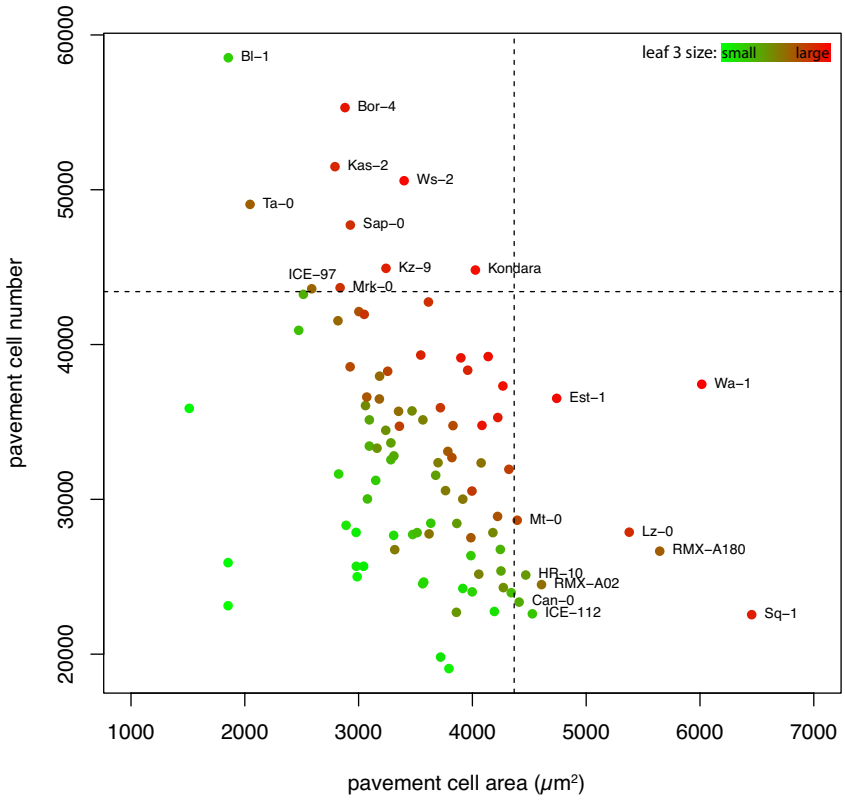


Figure 3: Pavement cell area and pavement cell number in relation to leaf 3 area at maturity. Dots are color coded according leaf 3 size under control conditions for all accessions, from small (green) to large (red). Dashed lines represent the 90% quantiles. Accessions that are above the 90% quantiles of pavement cell area or pavement cell number are labelled. Accessions with larger leaves have large cell numbers or cell areas or intermediate values for both. None of the accessions had both large and many pavement cells.

The dashed lines, representing the 90% quantiles, show that none of the accessions with the 10% largest pavement cell areas is present in the category having the 10% highest pavement cell numbers. From the 20 accessions within the category of either 10% largest pavement cells of 10% most pavement cells, 16 accessions have larger than average leaf sizes. Other accessions with larger leaves have both intermediate pavement cell number and areas. Pavement cell number and pavement cell area show a slight negative (r = -0.46) but significant correlation ($P <$ 0.001).

Under mild drought stress, the reduction in leaf 3 area correlates significantly with the reduction in pavement cell area (r = 0.53; $P < 0.0001$; Supplementary Figure 1K) and to a larger extent with the reduction in pavement cell number (r = 0.70; $P < 0.0001$ Supplementary Figure 1L). Although pavement cell area is on average slightly more reduced than cell number, the latter correlates better with the reduction in leaf 3 area at maturity. The three drought tolerant accessions (ICE-163, NFA-10, Pna-10) show a reduction in both pavement cell area and number that is lower than average. Remarkably Pna-10 even shows an increase in pavement cell number under mild drought conditions. The three sensitive accessions (Can-0, ICE-61, Rubezhnoe-1) show a reduction in both cellular parameters that is higher than average, however less pronounced for cell number (Supplementary Figure 2 A and B).

From the cellular analysis, it is clear that both cell division and expansion are affected by the mild drought stress and overall, in control and mild drought conditions, the pavement cell number shows a stronger relation to the final leaf size than the size of the pavement cells.

Effect of mild drought stress on leaf 3 during proliferation

The mild drought treatment was initiated 4 days after stratification (DAS) in order to affect the growth of leaf 3 during the proliferation stage. Leaves were therefore harvested at the last day that all leaf cells are dividing (full proliferation). The last day of full proliferation was determined for each accession by microscopic analysis as the last day before jigsaw-shaped epidermal cells started appearing at the leaf tip. This time point just before the transition between cell proliferation and cell expansion varied between 8 and 10 DAS, depending on the accession.

In control conditions, leaf 3 areas at the last day of full proliferation ranged from 0.009 mm^2 (SE = 0.0007 mm^2) in Ts-1 up to 0.149 mm^2 (SE = 0.009 mm^2)in Fei-0 (Figure 1C). To study the relation between leaf area at the last day of full proliferation and its mature size, the correlation between leaf size at the last day of full proliferation and at maturity was not significant (r = -0.04; $P = 0.7$). Also between the leaf size at proliferation and the pavement cell number at maturity, no correlation was found (r = -0.15; $P = 0.2$).

Already by the end of the cell proliferation phase in leaf development, the majority of the accessions are affected in their leaf 3 area by the mild drought stress. On average, the mild drought caused a reduction of 18% in leaf 3 area at this early developmental time point (Table 1) but we could not find any significant correlation between the leaf size reductions under mild drought at proliferation stage and maturity.

From this analysis, it was found that the leaves are clearly affected in size by the mild drought stress already at the proliferation phase. However, no clear relation was detected between the size of leaf 3 at proliferation phase and the size at maturity, nor with the cellular parameters at maturity.

Stomatal index

A major evolutionary adaptation of plants in response to fluctuations in water availability was the arising of stomata. These simple but effective structures allow the plant to regulate its temperature and its CO_2 uptake. Stomatal opening is tightly regulated in order to prevent unnecessary evapotranspiration (Kim et al., 2010). We analyzed whether mild drought stress altered the stomatal index in the 98 accessions, where the stomatal index is defined as the

percentage of stomata on the total cell number. Although the changes in stomatal index are significant over the 98 accessions, they are very limited. Under mild drought stress the stomatal index was on average 6% reduced (Table 1), however, the changes range from a reduction of 26% (Ler-1) to an increase of 13% (Wt-5). Under control conditions an average stomatal index of 45% is noted (Table 1; Figure 1F).

Genome wide association mapping (GWAS)

A genome wide association mapping was performed, in order to retrieve the SNPs and the genes that are associated with the phenotypes that were discussed above. Since phenotypes were measured in control and mild drought conditions and our main interest lays in the mild drought response of plants, we used a multi-trait mixed model (MTMM) to conduct the GWAS (Korte et al., 2012). This model allows for retrieving genetic loci that associate with the differential response upon mild drought stress of the growth-related phenotypes. The model describes the phenotype as a function of the genotype (G), environment (E) and the interaction of genotype and environment (G×E). As discussed in the introduction, population structure is a confounding factor and may lead to false positives in GWAS analyses (Vilhjálmsson and Nordborg, 2012). Therefore a kinship matrix was included as a random factor to correct for this. All the different growth-related phenotypes discussed above have been subjected to the MTMM. We focused specifically on the associations with the response to mild drought stress of these phenotypes.

Phenotype	Heritability - Control	Heritability - Mild Drought
Rosette Area	0.64 ± 0.13	0.91 ± 0.13
Leaf 3 at Maturity	0.87 ± 0.13	0.61 ± 0.18
Leaf 3 at Proliferation	0.22 ± 0.29	0.7 ± 0.31
Pavement Cell Area	0.45 ± 0.21	0.99 ± 0.01
Pavement Cell Number	0.99 ± 0.05	0.44 ± 0.31
Stomatal Index	0.69 ± 0.16	0.84 ± 0.13
Compactness	0.92 ± 0.06	0.99 ± 0.01

Table 2: Heritability estimates. Overview of the heritability estimates ± SE, for the different phenotypes in control and in mild drought conditions.

In order to detect genetic loci that can explain the variation in the phenotype, the dataset needs to contain a significant genetic factor that can explain the phenotypic variation better than unspecified factors (noise). Whether this is the case can be checked by calculating the heritability, which is the proportion of genetic variation on the total phenotypic variation that is present in the collection of accessions. Heritabilities were estimated for each phenotype in both control and mild drought conditions and were above 80% for all of the traits in at least one of both treatments except for leaf 3 area at proliferation (Table 2).

Associations with mild drought response of leaf growth-related phenotypes

Significant associations were only found for the differential response to mild drought stress of leaf 3 at the proliferation stage. The SNP's are located on chromosome 4 (Figure 4A) and within

the surrounding 10 kb* five genes are located: *NODULE-INCEPTION-LIKE PROTEIN 7* (*NLP7*), a pre-tRNA (AT4G24025), two genes encoding unknown proteins (AT4G24026, AT4G24030) and *TREHALASE1* (*TRE1*). However not significant, for the other phenotypes a number of SNPs displayed obvious peaks of association with the differential response to the mild drought stress.

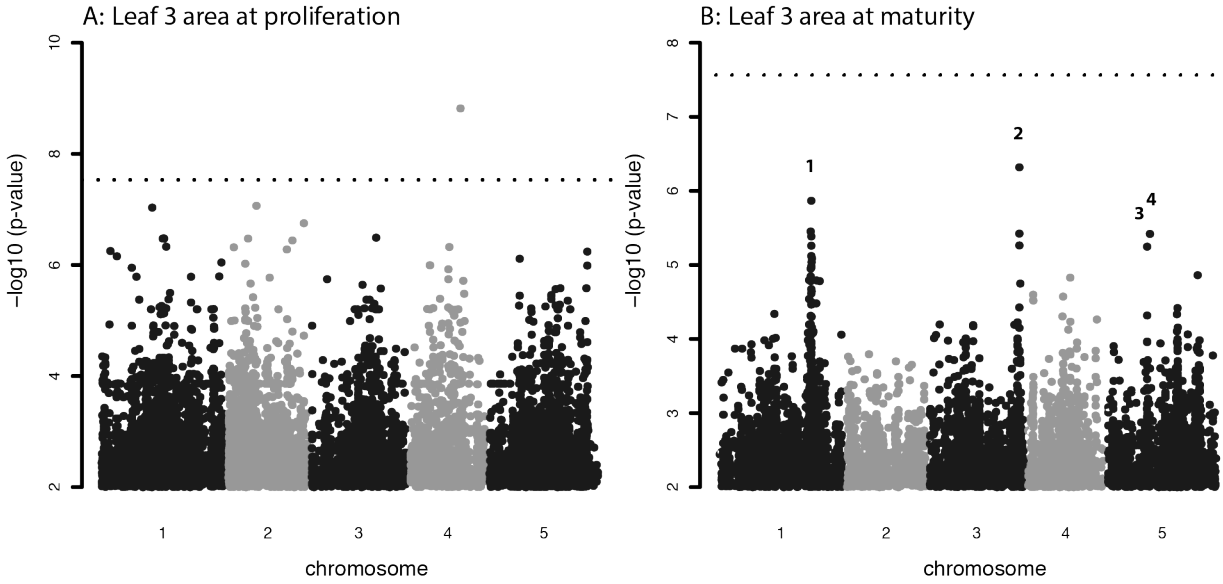


Figure 4: Manhattan plots showing significance of the associations of each SNP with the phenotype. A: Association of SNPs with the differential response of leaf 3 area at proliferation to mild drought stress. Dashed line indicates the Bonferroni corrected significance cut-off of 7.53 –log10 (p-value). B: Association of SNPs with the differential response of leaf 3 area at maturity to mild drought stress. Numbers indicate the peaks for which the genes were selected (**Supplementary Table 2**). Dashed line indicates the significance cut-off of 7.56 –log10 (p-value).

There was a clear increase in significance of the association with leaf 3 at maturity for 3 loci, in comparison to the background. 18 genes were linked to a clear peak on chromosome one, another seven genes were linked to the peak on chromosome 3 and five more genes linked to the two peaks on chromosome 5 (Figure 4B; Supplementary Table 2). One of these genes encodes miR171C, involved in controlling cell differentiation of the periphery of the shoot apical meristem (Schulze et al., 2010). Also for rosette area, pavement cell area and pavement cell number genes linked to peaks that were clearly distinguishable from background and had p-values below 10^{-4} were identified (Supplementary Figure 3).

Evidence from gene expression response to mild drought

The expression values of the genes associated to the mild drought response of the different growth related phenotypes were checked for their transcriptional response upon mild drought stress. The transcriptome data that was used is discussed below. The 23 genes that showed at least a two-fold expression difference between control and mild drought stress, averaged over the 89 sequenced accessions, are listed in Table 3.

---

* Linkage disequilibrium decays over 10kb in Arabidopsis (Kim et al., 2007). Therefore the causal SNP may be present within 10kb of an associated non-causal SNP.

Phenotype	Gene	Gene function
Leaf 3Area at Maturity	AT1G61667	unknown protein
Leaf 3 Area at Maturity	AT1G61680	TERPENE SYNTHASE 14 (TPS14)
Leaf 3 Area at Proliferation	AT4G24026	unknown protein
Leaf 3 area at Proliferation	AT4G24040	TREHALASE 1 (TRE1)
Pavement Cell Area	AT4G19950	unknown protein
Pavement Cell Area	AT4G19970	Nucleotide-diphospho-sugar transferase
Pavement Cell Number	AT1G30950	UNUSUAL FLORAL ORGANS (UFO)
Pavement Cell Number	AT1G52560	HSP20-like chaperones superfamily protein
Pavement Cell Number	AT2G19900	NADP-MALIC ENZYME 1 (NADP-ME1)
Pavement Cell Number	AT3G24110	Calcium-binding EF-hand family protein
Pavement Cell Number	AT5G14690	unknown protein
Rosette Area	AT1G61667	unknown protein
Rosette Area	AT1G61680	TERPENE SYNTHASE 14 (TPS14)
Rosette Area	AT3G24900	RECEPTOR LIKE PROTEIN 39 (RLP39)
Rosette Area	AT3G24982	RECEPTOR LIKE PROTEIN 40 (RLP40)
Rosette Area	AT3G28180	CELLULOSE-SYNTHASE LIKE C4 (CSLC4)
Rosette Area	AT3G28210	STRESS-ASSOCIATED PROTEIN 12 (SAP12)
Rosette Area	AT5G53710	unknown protein
Stomatal Index	AT1G35140	EXORDIUM LIKE 1 (EXL1)
Stomatal Index	AT1G60740	Thioredoxin superfamily protein
Stomatal Index	AT2G30930	unknown protein
Stomatal Index	AT3G43930	BRCT domain-containing DNA repair protein
Stomatal Index	AT5G44130	FASCICLIN-LIKE ARABINOGALACTAN PROTEIN 13 Precursor (FLA13)

Table 3: Genes detected in GWAS with differential expression upon mild drought. Genes are located within 10 kb of a SNP that associated to the mild drought response of the respective phenotype. The genes also show an average fold change higher than two, upon mild drought stress.

Two of the 23 genes were associated with the differential response of both rosette area and leaf 3 area at maturity. One of these genes encodes TERPENE SYNTHASE 14 (TPS14), involved in linalool synthesis (Tholl and Lee, 2011), the other gene encodes a protein with unknown function (AT1G61667). Five of the 23 genes associated only with rosette area: two receptor like proteins (RLP39 and RLP40), CELLULOSE SYNTHASE-LIKE C4 (CSLC4), the STRESS ASSOCIATED PROTEIN 12 (SAP12) and a gene of unknown function (AT5G53710). CSLC4 is involved in the biosynthesis of xyloglucan, a cell wall component (Chou et al., 2012). SAP12 is transcriptionally involved in the response to salt and cold stress (Ströher et al., 2009). Further, among the 23 genes there was a nucleotide-diphospho-sugar transferase (AT4G19970) and a gene of unknown function (AT4G19950) that associated with the differential response of pavement cell area. Similarly, for the mild drought response of pavement cell number, the flower and leaf development regulating *UNUSUAL FLORAL ORGANS* (*UFO*; Risseeuw et al., 2013), a heat-shock protein 20-like encoding gene (AT1G52560), the NADP-malic enzyme encoding *NADP-ME1*, a calcium-binding EF-hand family protein encoding gene (AT3G24110) and a gene of unknown function (AT5G14690) were retrieved. Finally, for stomatal index, *EXL1*, which controls growth and development under limited carbon and energy conditions (Schröder et al., 2011), a gene encoding thioredoxin (AT1G60740), a gene encoding the BRCT domain-

containing DNA repair protein (AT3G43930), a gene encoding a fasciclin-like arabinogalactan protein precursor (FLA13) and a gene with unknown function (AT2G30930) were among the 23 genes that showed a higher than two fold difference in expression between control and mild drought stress.

The associated genes were also screened for genes with a known differential expression response upon mild drought stress in young developing leaves, the common drought genes discussed in the previous chapter. Five associated genes overlapped with the 354 common drought genes. Three of the genes encode known proteins: a meprin and TRAF homology domain-containing gene (AT3G28220), an alcohol oxidase-related protein (AT4G19380) and the MYB transcription factor MATERNAL EFFECT EMBRYO ARREST 3 (MEE3) associated with the differential response of respectively the rosette area, pavement cell number and stomatal. Two of the overlapping genes had no known function and associated with rosette area (AT1G32460) and stomatal index (AT5G46790).

In conclusion, a number of SNPs clearly associated more significantly than others with certain phenotypes, however significance was reached for only few SNPs for the stress response of leaf 3 at proliferation. A plausible explanation for the low significance is the low statistical power due to the complexity of the traits and the limited number of accessions used.

Transcriptional changes in response to mild drought stress in the proliferating leaf of 89 accessions

To identify genes that are involved in the transcription response of young developing leaves to mild drought stress in the different accessions a genome wide transcriptome analysis was conducted by RNA-sequencing. Plants of 89 different accessions were exposed to mild drought stress from the moment the third leaf emerges from the shoot apical meristem (4 DAS). For each accession the third leaf was harvested for transcriptome profiling at the last day during which all epidermal cells are proliferating. Expression data were obtained for control and mild drought conditions for 22287 genes (RNA-sequencing data analysis is detailed in Material and Methods).

Overall, the differential expression upon mild drought stress differs strongly between the accessions. Only eleven 'signature' genes were found to show the same direction of differential expression (control versus mild drought) in at least 80 accessions (Figure 5; Supplementary Table 3).

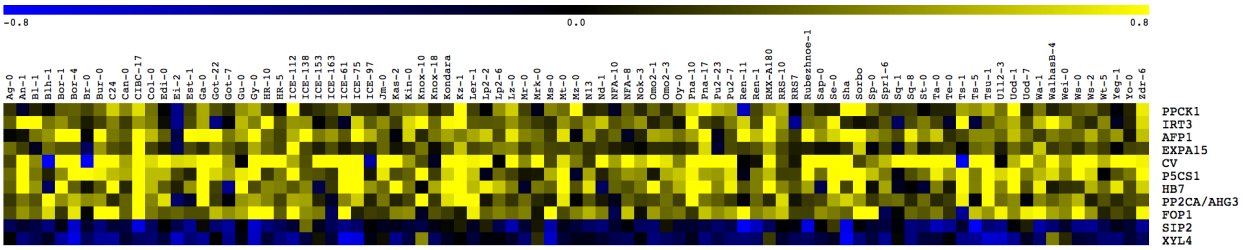


Figure 5: Differential expression of eleven 'signature' genes in 89 accessions. The genes show similar fold changes upon mild drought stress in at least 80 accessions. Fold changes are arcsin transformed (0.7 equals a 2 fold change).

The expression of these genes can be viewed as hallmarks for the transcription response to mild drought stress and are mainly involved in the classic drought response. ABI FIVE BINDING

PROTEIN (AFP1), PROTEIN PHOSPHATASE 2CA (PP2CA/AHG3) and HOME BOX 7 (HB7) are interactors of the ABA signaling pathway (Garcia et al., 2008; Rodrigues et al., 2013), from which HB7 has been described to regulate leaf growth and development upon drought (Söderman et al., 1996; Olsson et al., 2004; Valdés et al., 2012; Ré et al., 2014). ABA regulates proline synthesis by controlling the *P5CS1* expression in drought conditions (Strizhov et al., 1997; Verslues and Sharma, 2010). Proline functions as osmolyte, osmoprotectant and detoxifies reactive oxygen species (Szabados and Savouré, 2010). Expression of *SEED IMBIBITION 2* (*SIP2*), known to be involved in the synthesis of another osmoprotectant, raffinose (Saravitz et al., 1987; Taji et al., 2002; Castillo et al., 1990; Hincha et al., 2003) is however down-regulated. FOLDED PETAL 1 (FOP1) synthesizes wax esters (Takeda et al., 2013) that can contribute to cuticle thickening, an important adaptation to drought (Li et al., 2008; Kosma et al., 2009). Both EXPANSIN A15 (EXPA15) and BETA-D-XYLOSIDASE (XYL4) are involved in modifying the cell wall, which is important to control growth through expansion in drought (Wu and Cosgrove, 2000; Minic et al., 2004; Moore et al., 2008). Also regulation of primary metabolism is suggested by the up-regulation of PHOSPHOENOLPYRUVATE CARBOXYLASE KINASE 1 (PPCK1), which regulates the activity of the Krebs cycle enzyme phosphoenolpyruvate carboxylase and affects plant growth (Meimoun et al., 2009). CHLOROPLAST VESICULATION (CV) is a known stress-induced chloroplast disrupter (Wang and Blumwald, 2014) and its up-regulation suggest chloroplast breakdown upon mild drought stress. Remarkably a Zn/Fe transporter, IRON REGULATED TRANSPORTER 3 (IRT3*)* was up-regulated, but has to our knowledge no known function in the drought response. All eleven 'signature' genes, with the exception of the unknown gene AT2G25625, are part of the 354 common drought genes, described in the previous chapter.

A co-differential expression (co-expression of the differential expression) analysis was performed on all 22287 genes over the 89 accessions to detect potentially co-regulated and hence functionally related genes. Clusters of co-differentially expressed genes were obtained with the cluster affinity search technique (CAST; Ben-Dor et al., 1999). The CAST algorithm does not simply rely on the correlation between expression values of the different accessions, but it also considers the pattern of expression over the different accessions. Furthermore CAST does not require defining the number of clusters *a priori*, as is the case with k-means clustering. The CAST clustering resulted in 5447 clusters containing 1 to 2112 genes. In total 2208 clusters (41%) contained only one gene, meaning that these clusters did not show co-differential expression over all accessions with any other gene according this analysis. 385 clusters contained more than 5 genes, which cover in total 12869 genes. Of these 385 clusters, 50% contained between 9 and 5 genes.

Given the potential important roles in the mild drought response of young developing leaves for the above discussed 11 signature genes, the CAST clusters containing these genes were investigated in greater detail. The eleven genes were all part of different clusters with sizes ranging from three genes (cluster with AFP1) till 426 genes (cluster with EXPA15; Supplementary Table 4; Supplementary Figure 4-Supplementary Figure 14). The largest cluster (426 genes) with the cell wall modifier EXPA15, contained in total 56 genes that are involved in cell wall organization. The transcription factor HB7 clusters together with 41 other genes that were enriched for genes involved in the response to water deprivation and the response to ABA

(GO-enrichment: $P > 0.001$; resp. 15 and 16 genes). This fits with the regulatory role of HB7 on ABA signaling (Valdés et al., 2012) and its involvement in the regulation of growth upon drought (Söderman et al., 1996; Olsson et al., 2004; Ré et al., 2014). In the cluster with *XYL4*, three of the nine genes were also involved in cell wall modifications: *XYL4*, *XYL2* and *PLEIOTROPIC DRUG RESISTANCE 7* (*PDR7*). Three of the 11 signature genes (*PPCK1*, *P5CS1*, *PP2CA/AHG3*) were present in the same co-differential expression cluster, which was enriched for genes in the hyperosmotic response (GO-enrichment: $P = 0.002$); *P5CS1*, *PP2CA/AHG3*, *CBL-INTERACTING PROTEIN KINASE 6* (*CIPK6*) and a mannose-binding protein (AT1G52000). The above listed examples show that the co-expression clusters obtained with the CAST algorithm may represent functional clusters that are putatively co-regulated.

In conclusion, there is substantial variation in gene expression between the accessions. Only eleven genes showed similar differential expression in at least 80 accessions. With ten of the eleven genes previously found to be involved in a conserved response to mild drought, they are potentially important players in the mild drought response of young developing leaves. The variation in differential expression did show a pattern over the different accessions for the majority of the genes and were co-differentially expressed in clusters of variable sizes. This clustering hints for co-regulation and functional grouping of genes.

Stress predictors

The transcriptome data allowed for a modeling approach to learn which set of genes could predict whether a sample was subjected to control or mild drought based on the expression values. A classification model was built to distinguish between the samples in the two treatment groups: control and mild drought stress. In other words, the setup allowed for a two-class classification where samples under stress were seen as the positive class while the control samples were set as the negative class. In a first step, to reduce the large gene space of 22287 filtered genes (detailed in Material and Methods) and thus decrease the complexity of the classification model, a one-way ANOVA with the stress condition as a factor was performed before building the classification model. After correcting for multiple hypothesis testing, genes were retained based on a significance threshold of 0.05. This delivered a set of 283 genes capable of predicting whether a sample is exposed to mild drought or control conditions in young developing leaves. In a second step, a classification model was built based on this set of genes. The classification was performed by support vector machines (further details in Material and Methods) and resulted in an F-score of 0.859. The F-score is the harmonic mean between sensitivity and precision, which can vary from 0 to 1. An F-score of 1 is retrieved in case of maximum sensitivity and precision. Another performance measure, the area under the curve (AUC) was 0.926, where an AUC value of 1 implies a perfect separation of mild drought and control samples, while 0.5 implies random classification (further detailed in Material and Methods). Both performance indicators point out that the model was highly successful in separating control from mild drought treated samples.

Biological function of the stress predictors

To determine a functional bias in the selected genes, we performed a GO enrichment analysis. The list of 283 stress predictors is highly enriched in genes responding to water deprivation,

osmotic stress and abscisic acid (Table 4). Besides these stress related functional enrichment, also 42 genes are involved in regulation of gene expression, which was a significantly enriched GO category ($P = 0.04$; 0.4 log2 enrichment). Moreover, the stress predictors are enriched ($P < 0.001$; 4.1 log2 enrichment) for the common drought genes identified by Clauw et al. (2015). In total 79 of the 354 common drought genes were also found to be stress predictors (Supplementary Table 5). Among these 79 genes we found 18 ABA-related genes from which 10 genes encoded components of the core signaling pathway: ABRE-BINDING FACTOR 3 (ABF3), PROTEIN PHOSPHATASE 2CA (PP2CA/AHG3), ABA INSENSITIVE 1 (ABI1), ABI2, HYPERSENSITIVE TO ABA 1 (HAB1), HIGHLY ABA-INDUCED PP2C GENE (HAI1), HAI2, PYRABACTIN RESISTANCE 1-LIKE 1 (PYL1), PYL4 and PYL6. Two of the ABA-responsive genes are the proline metabolizing DELTA1-PYRROLINE-5-CARBOXYLATE SYNTHASE 1 (P5CS1) and PROLINE DEHYDROGENASE 1 (PDH1/ERD5). Furthermore, 8 genes involved in cell wall modification were present: *BETA-XYLOSIDASE 2* (*BXL2*), *EXPANSIN A3* (*EXPA3*), *EXPA15*, *EXTENSIN 3* (*EXT3*), pectin-lyase like encoding genes (AT1G48100, AT1G10640, AT1G67750) and *BETA-D-XYLOSIDASE 4* (*XYL4*). Also the eleven 'signature' genes with similar expression responses in at least 80 accessions were identified as stress predictors. Within the common drought genes a subset of 87 genes was identified to be specific for the response of young developing leaves to mild drought stress (Clauw et al., 2015). Thirteen of these 87 specific mild drought responsive genes of young developing leaves are found back among the stress predictors (Supplementary Table 5). Furthermore, the 204 stress predictors that do not overlap with the common drought genes contain 18 genes associated to responses to water deprivation (GO enrichment: $P < 0.001$; 1.95 log2 enrichment; Supplementary Table 6). Among these 18 genes the ABA biosynthesis rate-limiting enzyme encoding *NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 3* (*NCED3*) is found together with *RESPONSIVE TO ABA 18* (*RAB18*) and *HISTONE H1-3* (*HIS1-3*), which are known ABA responsive genes. Also three bHLHs (*bHLH100*, *bHLH38* and *bHLH39*) and the GA receptor *GA INSENSITIVE DWARF 1B* (*GID1B*) were among these 18 genes. Among the stress predictors there are also two GA biosynthesis genes (*GA20OX1* and *GA20OX2*). GA is involved in the growth regulation upon environmental perturbations (Claeys et al., 2014). The above analysis of the biological functions of the stress predictors shows that the stress predictors are clearly enriched for genes that have a function in drought responses, it can hence be hypothesized that the stress predictors with no known role in drought are also active in the transcriptional response to mild drought stress and are interesting candidates for further investigation.

Description	GO-term	Log2-Enrichment	Q-value
response to water deprivation	GO:0009414	2.43	3.52E-15
response to water	GO:0009415	2.42	5.10E-15
response to superoxide	GO:0000303	4.43	7.80E-11
response to inorganic substance	GO:0010035	1.44	1.29E-10
response to abscisic acid	GO:0009737	1.8	1.30E-09
hyperosmotic salinity response	GO:0042538	3.03	1.53E-09
response to oxygen-containing compound	GO:1901700	1.07	2.93E-09
response to lipid	GO:0033993	1.61	3.73E-09

single-organism cellular process	GO:0044763	0.51	4.41E-09
response to alcohol	GO:0097305	1.64	1.08E-08
cellular response to abscisic acid stimulus	GO:0071215	2.23	1.52E-08
response to chemical	GO:0042221	0.82	2.88E-08
single-organism process	GO:0044699	0.4	3.03E-08
hyperosmotic response	GO:0006972	2.55	4.80E-08
cellular response to lipid	GO:0071396	1.91	1.50E-07
cellular response to oxygen-containing compound	GO:1901701	1.42	1.53E-07
cellular response to alcohol	GO:0097306	2.02	1.61E-07
abscisic acid-activated signaling pathway	GO:0009738	2.15	1.76E-07
response to stimulus	GO:0050896	0.55	3.27E-07
response to stress	GO:0006950	0.72	3.54E-07

Table 4: 20 most significantly enriched GO categories in the 283 stress predictors. The Q-values are Bonferroni corrected p-values.

Co-differential expression

To find out which of the stress predictors show similar transcriptional responses to mild drought stress over the 89 accessions, a co-differential expression network was built (Figure 6). Therefore the expression fold change between control and mild drought stress was taken for the 283 stress predictors. Subsequently the Pearson correlation was calculated for each of the possible gene pairs over the 89 accessions. This resulted in a network of 63 genes that showed high connectivity. One, small subnetwork contains *bHLH38*, *bHLH39*, *bHLH100* and AT2G14247, encoding an unknown protein. All three bHLHs are involved in leaf cell differentiation and chloroplast development (Andriankaja et al., 2014) and showed a Pearson correlation of 0.8.

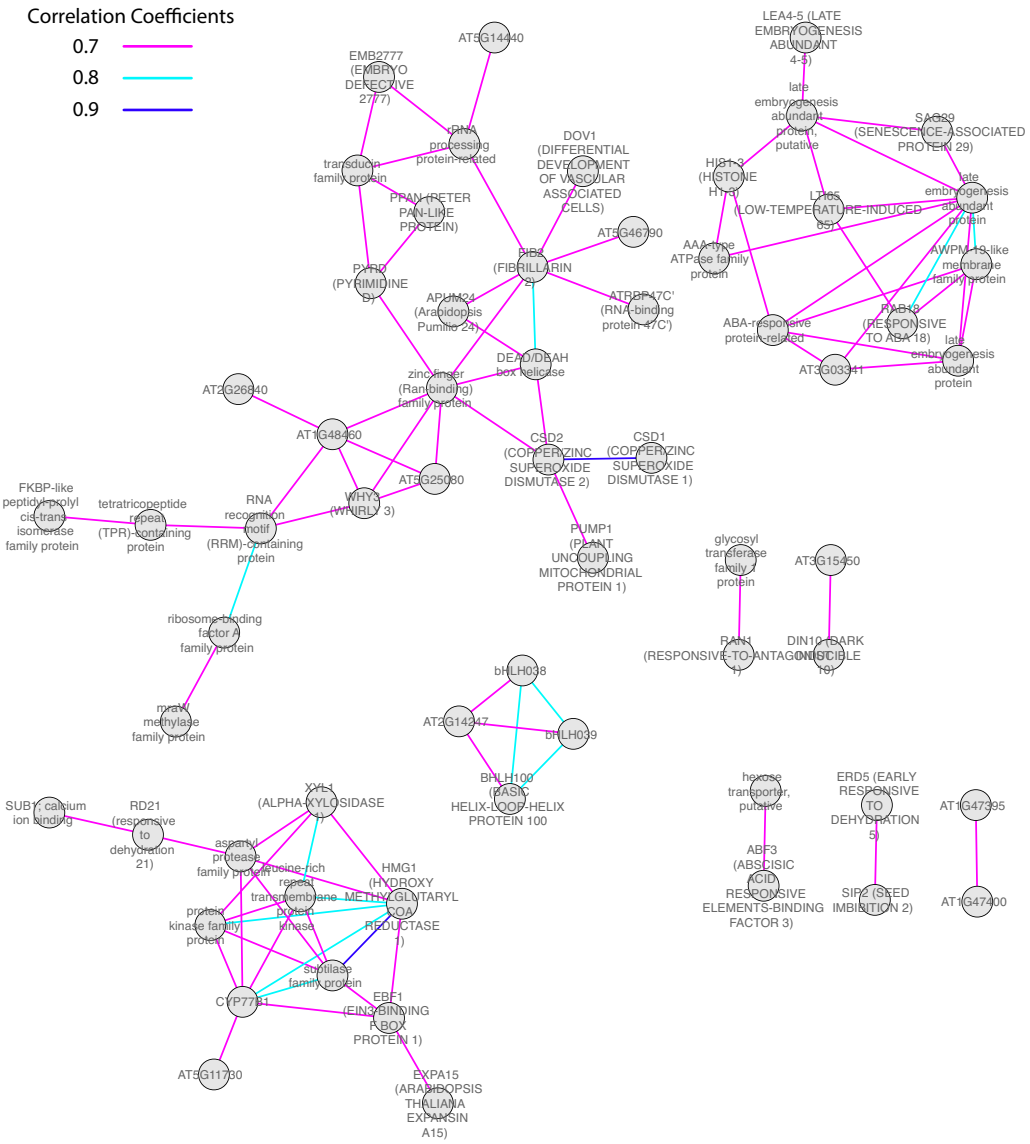


Figure 6: Co-differential expression network of the stress predictors. Network visualization of the Pearson correlation of the differential expression upon mild drought stress between the different stress predictors over 89 accessions. Central in the figure is the cluster located with bHLH38; bHLH39 and bHLH100.

In conclusion, the stress predictors are enriched for genes and processes involved in drought responses and for genes that were previously found to be involved in proliferating leaf tissue under mild drought conditions. The stress predictors are partly co-differentially expressed under mild drought stress and potentially contain key players in the mild drought response.

eGWAS

The phenotype of a plant is defined by its surrounding environment and the genotype. The genotype influences the phenotype through the transcription of genes that are translated into proteins, which in turn will have an active function in shaping the plant's phenotype. Gene transcription and protein activity is not purely regulated by the genotype but is also influenced by the environment. On the other hand, the environment also influences gene transcription in a genotype-dependent manner, and hence the interaction of both genotype and environment will have a large influence on the phenotype.

As gene transcription is an important intermediate step between the genotype and the phenotype, we wanted to identify the genetic loci that influence transcription in young developing leaves and which ones are doing this in interaction with an environmental perturbation, in particular mild drought stress. To do so, the transcriptome data of young developing leaves from 89 accessions, grown in control and mild drought stress, were associated to their respective genotypes in a GWAS approach. In order to retrieve genetic loci that associate with the differential expression upon mild drought stress the multi-trait mixed model (MTMM) was used (Korte et al., 2012). The same model was used for the GWAS analysis of the mild drought responses of the growth-related phenotypes, but here the phenotype is gene expression. Therefore we refer to this kind of analysis as expression genome wide association mapping or eGWAS. The MTMM was run independently for 3393 genes that were selected as the genes where, in this dataset, the genetic factors could explain the expression values much better than unspecified factors (noise). In order to perform this selection, the heritability was estimated as the proportion of the genetic variation on the total expression variation that was present in the collection of accessions. The genes with a heritability of more than 95% for their expression in control or mild drought conditions were selected for the eGWAS analysis.

The MTMM allows for testing the effect of a SNP on the expression of a particular gene independent of the treatment (both control and mild drought conditions), referred to as the common test. On the other hand, also the effect of a SNP on the differential expression upon mild drought stress can be evaluated, further referred to as the trait specific test, which evaluates G×E.

The MTMM converged for 1715 of the 3393 genes. For the genes where the MTMM did not converge, the iterative process was not able to find appropriate estimates for the MTMM. Hence, the associations found with these models are not meaningful and were thus not included in the further analysis. For numerous genes we observed a great number of loci that associate with the expression of one gene, visible as horizontal bands when plotting the position of the associated SNPs against the position of the gene they potentially regulate (Figure 7 A and B), which is most likely an indication of model inflation (See Box 1).

In order to exclude genes for which the MTMM was inflated, a first selection was done for genes with a genomic inflation factor (λ) equal to 1 ± 0.1, where 1 indicates that there is no inflation. This selection resulted in 1436 genes that showed no inflation according the λ values. However, not all genes with inflated models could be discarded with this selection, as can be seen in Figure 7 (C and D). To exclude also the remaining genes suffering from model inflation, a more stringent selection was performed based on the normality of the gene expression distribution by using the D'Agostino test (See Box 1).

Box 1: Model inflation, causes, consequences and solutions.

Model inflation is, in this particular case, the situation where an extensive number of loci (several hundreds or more) associate with the gene expression values. It is biologically unlikely that so many loci regulate the expression of a gene, most probably a large part of the associations are false positives. In statistical terms, there are much more significant p-values than would be expected, a situation which is called model inflation. This situation may arise when the distribution of the gene expression values of the different accessions for a particular gene is not normal and coincides with the genetic relatedness of the accessions: e.g. gene X is strongly expressed in group A of accessions and not expressed in group B. The separation of accessions in groups A and B coincides with the population structure of the collection. Accessions in group A are thus genetically strongly related, the same for accessions in group B and very few alleles are shared between both groups. Because there is a clear separation between groups A and B in gene expression but also in the alleles the accessions carry, there will be a strong association between gene expression levels and all the SNPs for which the alleles are separated between the two groups. The higher the relatedness within each group, the more SNPs for which this will be the case.

The association of each separate SNP with the expression value is judged on its significance, expressed as a p-value. Due to the large number of SNPs (±4,000,000), a normal distribution can be expected for the p-values. When the model is inflated however, there will be more significant p-values than expected, hence the distribution will be skewed and the median of the distribution will be shifted. To detect a skewed distribution of the p-values and thus whether the model is inflated, the genomic inflation factor (λ) is calculated as the ratio between the median of the observed p-value distribution and the expected median (0.5 for p-values). A λ-value of 1 indicates that observed and expected median are equal, deviations from 1 indicate a shift in the median of the p-values and the extent to which the model is inflated and thus the extent of false positive associations (Devlin and Roeder, 1999).

Model inflation may be caused by non-normal distributed expression values (as discussed above). The distribution characteristics of the expression values over the accessions can thus be used as an indicator for model inflation. The normality of a distribution can be described by the skewness (how symmetric the distribution is; expected value is 0) and the kurtosis (how sharp the peak of the distribution is; expected value is 0). The D'Agostino test takes both parameters into account to test whether the distribution is significantly different from the normal distribution (D'Agostino, 1986).

Genes that showed non-significant ($P > 0.05$) difference from the normal distribution according the D'Agostino test were selected for further analysis. Combining the selection based on convergence, genomic inflation factor and the D' Agostino test resulted in a final set of 371 genes for which the eGWAS results were further analyzed (Figure 7 E and F). From the eGWAS analysis, associated SNPs for both treatment independent and treatment dependent expression were detected in the common and trait-specific test respectively. Results of both tests for the 371 genes are discussed below.

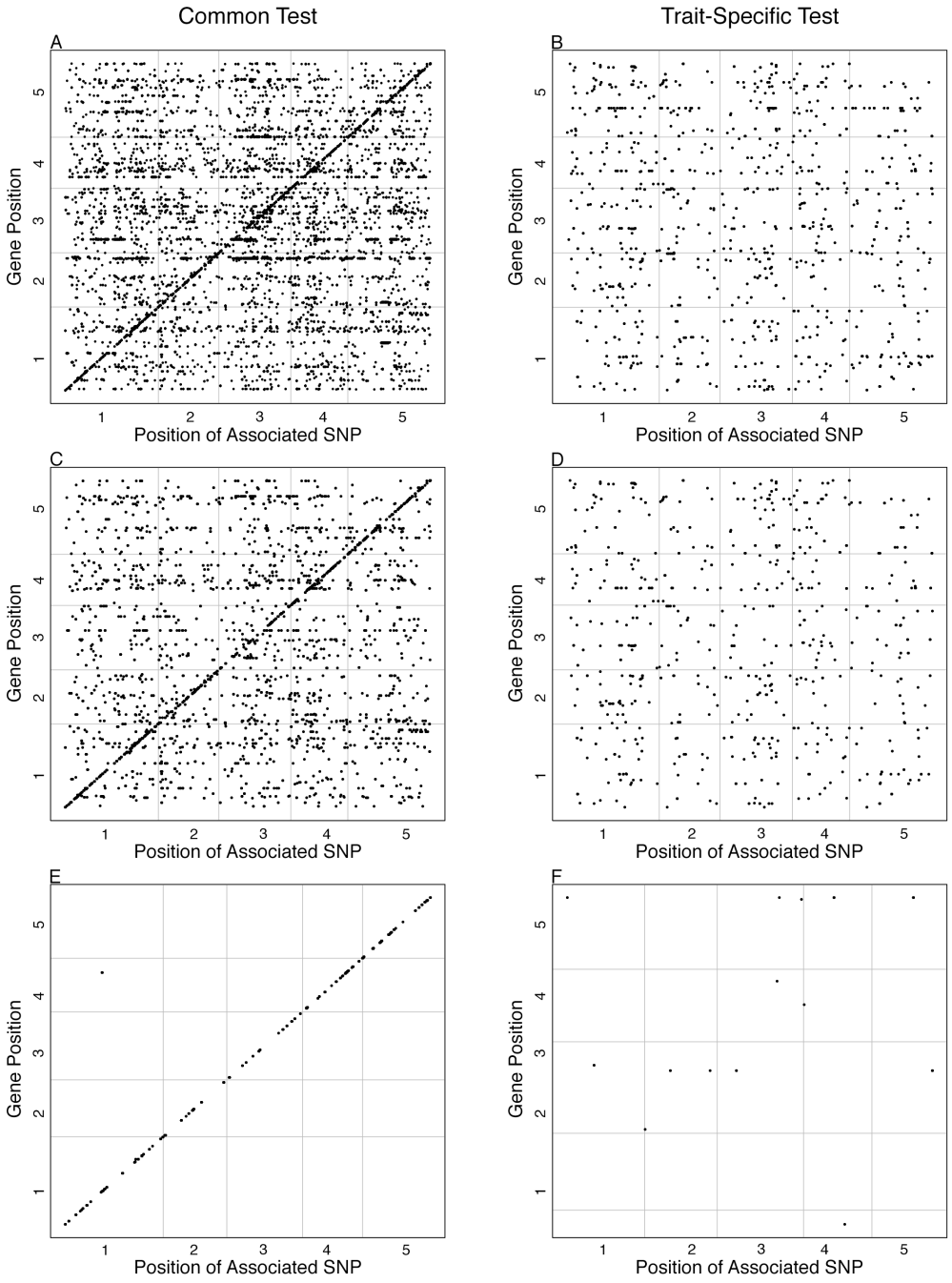


Figure 7: Overview of eGWAS selection. The different scatterplots show for each of the genes (positions on y-axes) the SNPs that associate with treatment independent expression (common test: A, C, E) and SNPs that associate with treatment dependent expression (trait-specific test: B, D, F). The different panels show the results of the stepwise selection procedure. A and B are the results of the 1715 genes for which the model converged. C and D show the results of the 1436 genes selected on model convergence and the genomic inflation factor (λ). E and F show the results of the 371 genes selected on model convergence, genomic inflation factor (λ) and the D'Agostino test.

Regulators of treatment-independent gene expression - common test

Of the 371 analyzed genes, 95 had at least one SNP that associated with the variation in treatment-independent gene expression, as detected with the common test. In Figure 7E showing the location of the 95 genes and the SNPs associated with their treatment-independent expression, a clear diagonal is visible. This diagonal band contains SNPs that are located in the

proximity of the gene whose expression they are associated with. This band contains potential *cis* regulatory SNPs for 79 genes. 34% of the associated SNPs are located within 1kb of the transcribed region (as determined by the latest Arabidopsis genome annotation, TAIR10; www.arabidopsis.org) of the regulated gene and a region of 15kb up- and downstream the gene covers 80% of the associated SNPs (Supplementary Figure 15). Despite this spread, the largest concentration of SNPs is located close to the gene. From the spatial distribution of the SNPs identified in the common test we observed that the proportion of SNPs clearly increased in the promoter region 2kb upstream of the transcription start site, with the largest density of associated SNPs located within the 1kb upstream region, after which the proportion of SNPs rapidly declines (Figure 8).

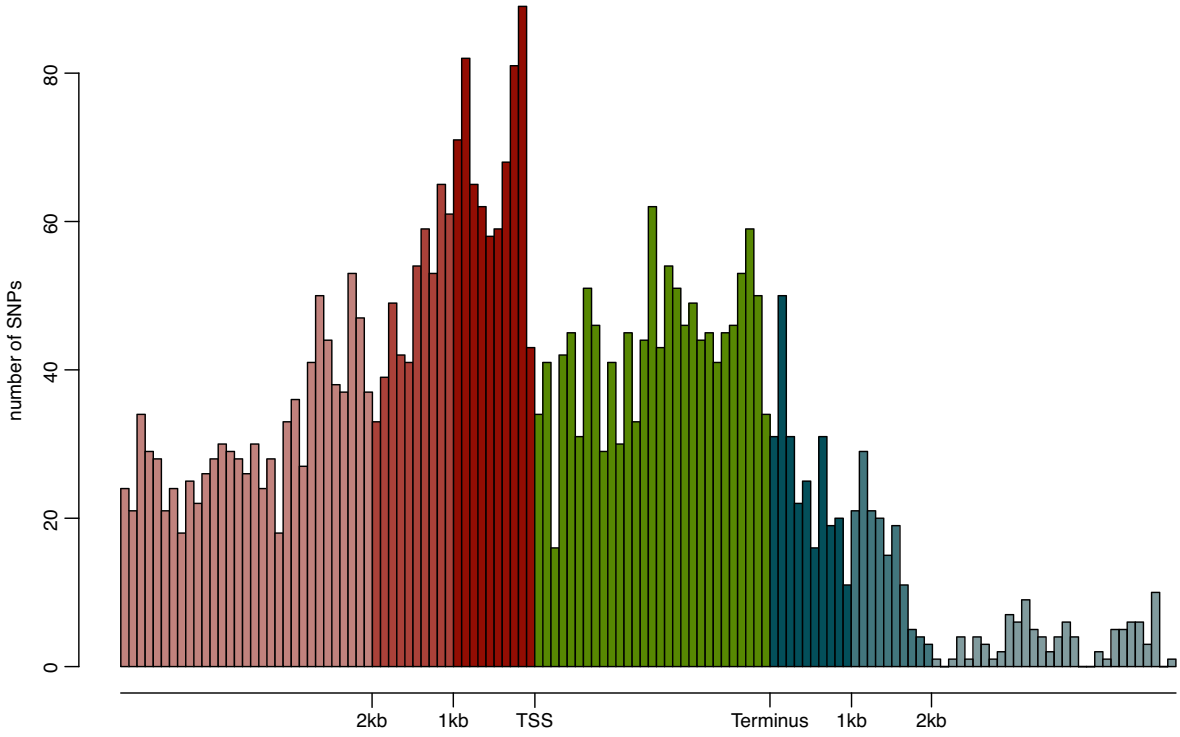


Figure 8: histogram of locations of SNPs associated to treatment independent expression. The transcription start site (TSS) and stop site (Terminus) are the URT boundaries as determined by the latest Arabidopsis genome annotation (TAIR10; www.arabidopsis.org). Indicated in red are the associated SNPs located upstream of the TSS. The 1 kb, 2 kb and SNPs further upstream are colored in different shades of red. In blue are the associated SNPs downstream of the transcription stop site. Different shades of blue indicate the different distances downstream of the transcription stop site (1kb, 2kb and further). In green are the associated SNPs located in the gene itself, mapped relative to the average gene length (± 300 bp).

Nevertheless, associated SNPs are also located in the transcribed region. These may point toward regulation by alternative splicing, certainly when the SNPs are located in or around introns. Alternative splicing occurs in approximately 60% of the intron-containing genes and is likely to play a role in plant growth, development and responses to environmental influences (Staiger and Brown, 2013). Downstream of the transcribed region, the number of SNPs that regulate expression rapidly declines. In the subsequent part of the thesis *cis* is defined as the region within 1kb up- and downstream of the transcribed region, respectively including the gene body.

To detect the effect of each of the *cis* SNPs on gene expression, the accessions were split in two haplogroups according the allele of the SNP. The expression fold change of the regulated gene between the two haplogroups was calculated, both for control and mild drought stress. The fold

changes of gene expression between the haplogroups varied from 1.2 (0.26 log2 fold change) to 128 (7 log2 fold change), with 71% of the SNPs showing a fold change of at least two (Figure 9A).

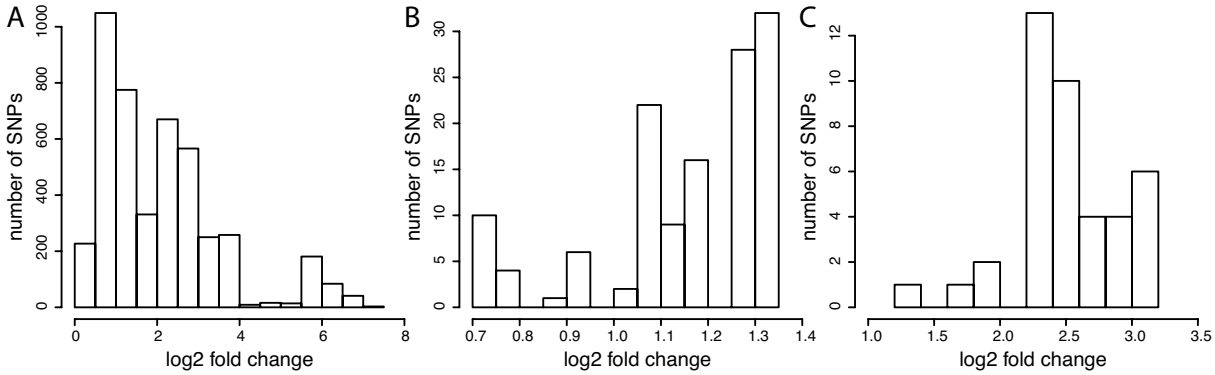


Figure 9: Effect on gene expression of associated SNPs. Histograms show the expression differences (log_2 fold change) between the haplogroups of the associated SNPs. A; Expression differences caused by *cis* SNPs associated with treatment independent expression. B: Expression differences caused by *trans* SNPs associated with treatment independent expression. C. Expression differences caused by *trans* SNPs associated with treatment dependent expression.

In addition to the *cis* located SNPs, 75 genes associated with SNPs located outside the 1kb surrounding of the gene. For 59 of the 75 genes both *cis* and *trans* loci were detected. The SNPs located in *cis* and *trans* are, over all genes together, distributed in a continuum (Supplementary Figure 15). For three genes, the SNPs were clearly separated from this continuous distribution and were located over 100kb from, or on another chromosome than, the gene they regulate. These three genes are further discussed as examples of *trans* regulated genes of the common test. The three genes encode proteins with different functions: the S-domain receptor kinase S-DOMAIN1-29 (SD1-29); AT4G27050 encodes an F-Box/RNI-like protein; the TRYPTOPHAN SYNTHASE BETA-SUBUNIT2 (TSB2), catalyzing the last step in tryptophan biosynthesis. The fold changes of the expression values between the haplogroups of the associated SNPs in trans ranges from 1.6 (0.70 log2) to 2.5 (1.32 log2; Figure 9B).

Ten SNPs were found to be associated with the expression of SD1-29. More specifically, 8 SNPs were located in the AT1G59660 gene and the 2 other SNPs in two transposable elements (AT1TE72360, AT1TE72365) downstream of AT1G59660 (Figure 10A). The AT1G59660 gene encodes a nucleoporin autopeptidase, which is part of the nucleopore complex (NPC). The NPC mediates the transport from the nucleus to the cytoplasm of a.o. mRNA. The nucleoporins that constitute the NPC were shown in Drosophila to be involved in the transcriptional regulation developmental and cell cycle genes (Capelson et al., 2010; Tamura et al., 2010). Due to linkage disequilibrium, the causal SNP/gene might however still be located within 10 kb of the associated SNPs. Especially three F-box proteins (AT1G59675, AT1G59680, AT1G59690), in this region, are good candidates to be potential regulators, but also the two genes with unknown function (AT1G59650, AT1G59710) may have a regulatory function.

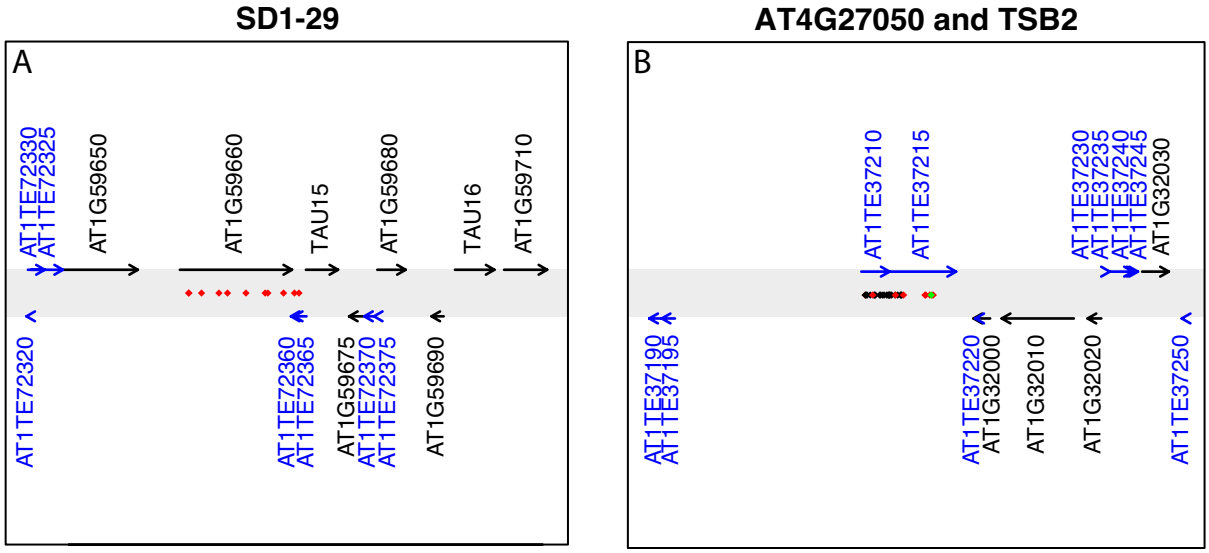


Figure 10: Schematic overview of associated SNP regions. Location of the SNPs (diamonds) associated with treatment independent expression of SD1-29 (A), and AT4G27050 and AT4G27070 (B). Arrows indicate the direction of transcription, for genes in black and for transposable elements in blue. Red diamonds indicate the SNP locations (B: black diamonds are associated to expression of both AT4G27050 and AT4G27070; red diamonds are specific for AT4G27050; green diamonds are specific for AT4G27070).

For AT4G27050 and TSB2 (AT4G27070), respectively 31 and 24 SNPs located on chromosome 1 were associated to their respective expression. Both genes are located within 10kb of each other and are separated by only one gene (AT4G27060). Remarkably, the associated SNPs of both genes are located within the same 10 kb region. Moreover, 23 of the SNPs associate with expression of both genes (Figure 10 B). The SNPs clearly cluster in a region that contains two transposable elements (AT1TE37215 and AT1TE37210). Within 10kb of the associated SNPs 4 genes are found, which encode a myosin heavy chain-related protein (AT1G32010), an F-box family protein (AT1G32020) and two unknown proteins (AT1G32000, AT1G32030).

Regulators of differential gene expression upon mild drought - trait specific test

From the trait-specific test of the multi-trait mixed model, SNPs associated with differential expression upon mild drought were found for eight genes. The eight genes encode for the MAJOR LATEX PROTEIN-LIKE PROTEIN 28 (MLP28), a plasma-membrane choline transporter (AT3G03700), an unknown protein (AT3G44430), PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE C8 (PLC8), a disease resistance protein (AT4G16900), SUCCINATE DEHYDROGENASE3-2 (SDH3-2), a FAD-binding berberine family protein (AT5G44410) and a protein kinase (AT5G45430). The AT5G44410 gene is involved in cell wall dynamics (Irshad et al., 2008) and SDH3-2 is targeted to the mitochondria where it is involved in both the citric acid cycle and the electron transport chain, and can be linked to growth. However, none of the above genes have to our knowledge been described as being involved in drought responses in plants or any other abiotic stress.

For five of the eight genes (SDH3-2, MLP28, AT3G03700, AT4G16900, AT5G44410) a single SNP was associated with differential expression upon mild drought stress. For PLC8, two associated SNPs were found, located within 5 kb of each other. Differential expression of AT3G44430 showed 12 associated SNPs spread over four different regions on chromosomes 2,

3 and 5. For AT5G45430, 22 associated SNPs were found, spread over 3 regions on chromosomes 1, 3 and 4. While the majority of the associated SNPs determined in the common test were located in *cis*, all of the trait-specific associated SNPs were located in *trans*. This suggests an important role for *trans* regulation in controlling expression levels upon environmental perturbations. To detect the effect of each *trans* SNP on differential gene expression, the fold change of the differential expression between the two haplogroups was calculated. The effect of the associated SNPs differed from a 2.5 up till an eight-fold change (resp. 1.32 and 3 log2 fold change; Figure 9C) in differential expression upon mild drought stress. All but one SNP resulted in a fold change of at least 2.8 (1.5 log2 fold change). The effect on differential expression of the *trans* regulators was much greater than for the *trans* regulators of the treatment-independent expression. Regulators in *trans* are thus likely to play an important role in regulating gene expression upon environmental perturbations.

Potential regulator genes of differential expression upon mild drought stress

The SNPs that were found to be associated with differential expression upon mild drought stress were located in varying types of genetic elements such as protein-encoding genes, transposable elements (TE) or in non-annotated sequences. Here, the different potential regulators that are linked to the associated SNPs will be discussed.

For six of the eight genes (MLP28, AT3G44430, AT4G16900, SDH3-2, AT5G44410, AT5G45430) the associated trait-specific SNPs were located within a gene body (Figure 11). The differential expression of MLP28 is associated with a SNP located in *MALE-STERILE 5/THREE DIVISION MUTANT 1* (*MS5/TDM1*; Figure 11A). The exact function of MLP28 is currently unknown, but is suggested to be involved in the root gravitropic response (Kimbrough et al., 2004). The potential regulator *MS5/TDM1* gene is one of the key players in regulating meiotic progression by inhibiting CDK activity (Bulankova et al., 2010; Cromer et al., 2012), however its exact function is unknown. In the same region also *HEN1* and the highly similar (75% sequence similarity) adjacent gene AT4G20920 are located. HEN1 is a general stabilizer of miRNAs (Ji and Chen, 2012; Baranauskė et al., 2015) and has been shown to be a rate limiting factor in the specific accumulation of miR157d and miR319/JAGGED AND WAVY (JAW; Tsai et al., 2014). JAW in turn is known to affect leaf development through microRNA mediated cleavage of specific members of the TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR1 (TCP) class of transcription factors (Palatnik et al., 2003). The TCP gene family of transcription factors are involved in plant developmental processes and specifically in cell proliferation (Martín-Trillo and Cubas, 2010). Thus two potential regulators of MLP28 are related to cell division, of which HEN1 ahs a link with mitotic cell division. Through the regulation of JAW and subsequently the TCP transcription factors, HEN1 can indirectly regulate gene expression. However the link to MLP28 and its role in young developing leaves in mild drought stress needs further investigation.

For the differential expression of the unknown protein AT3G44430, associated SNPs were detected in four different regions spread on chromosome 2, 3 and 5. The SNPs located on the fifth chromosome were clustered around and within TCP6 (AT5G41030) (Figure 11B). Other genes in within 10kb of the associated SNPs encode a MYB-transcription factor (AT5G41020),

an RNA-polymerase (RPB12) and ALIPHATIC SUBERIN FERULOYL-TRANSFERASE (ASFT).

The disease resistance gene AT4G16900 shares 90% sequence identity with its neighboring R-genes (AT4G16890, AT4G16950, AT4G16990). The sole SNP associated to differential expression of AT4G16900 is located in a DNA-binding storekeeper protein-related transcriptional regulator (AT4G00390; Figure 11 C), a class of transcriptional factors identified in potato (Zourelidou et al., 2002) of which further functional details are unknown. Five kb downstream of the SNP, the METHYL-CPG BINDING DOMAIN 3 (MBD3) gene is located, which is linked to epigenetic regulation of gene silencing by cytosine methylation (Zemach and Grafi, 2003). Upstream of the DNA-binding storekeeper gene, the *FACTOR OF DNA METHYLATION2/INVOLVED IN DE-NOVO 2 (IDN2) PARALOG 2/IDN2-LIKE 2* (*FDM2/IDP2/INDL2*) gene is located, which is involved in RNA-directed methylation of cytosine (Ausin et al., 2012; Zhang et al., 2012). The cluster of R-genes has been shown to be epigenetically regulated (Stokes et al., 2002). Despite the transcription regulatory function of AT4G00390, it cannot be excluded and may even be more plausible that the methylation genes MBD3 and FDM2/IDP2/INDL2 are regulating expression of the disease resistance gene AT4G16900.

The SDH3-2 gene encodes one of the two membrane proteins that anchor the succinate:ubiquinone oxidoreductase complex in mitochondria. This complex, which is also referred to as complex II is functional in both the citric acid cycle and the aerobic respiratory chain where it catalyzes oxidations of succinate into fumarate and reduces ubiquinone to ubiquinol (Figueroa et al., 2002). The differential expression of SDH3-2 was associated to one SNP located in *PHYTOSULPHOKINE 3 PRECURSOR* (*PSK3*; Figure 11D), a peptide hormone known to be involved in plant growth and widely expressed in leaves, cotyledons and root tips. As one of the 5 PSK precursors, PSK3 is involved in the hormonal regulation of cell longevity and cell proliferation (Matsubayashi et al., 2006). Although, the regulatory mechanism of the PSK signaling pathway still needs to be unraveled. Recently it has been shown that PSK is capable of altering transcript levels and is involved in stem cell replenishment in the root tip (Heyman et al., 2013). In the ornamental plant *Zinnia elegans*, PSK regulates the expression of stress related genes (Motose et al., 2009). Another potential transcriptional regulator within 10kb of the associated SNP is *HISTONE DEACYTELASE2A/HISTONE DEACYTELASE3* (*HD2A/HDA3*). By removing acyl groups from the histones, histone deacytelases induce a tighter wrapping of DNA around the histones and hence reduce availability for transcription of those wrapped regions.

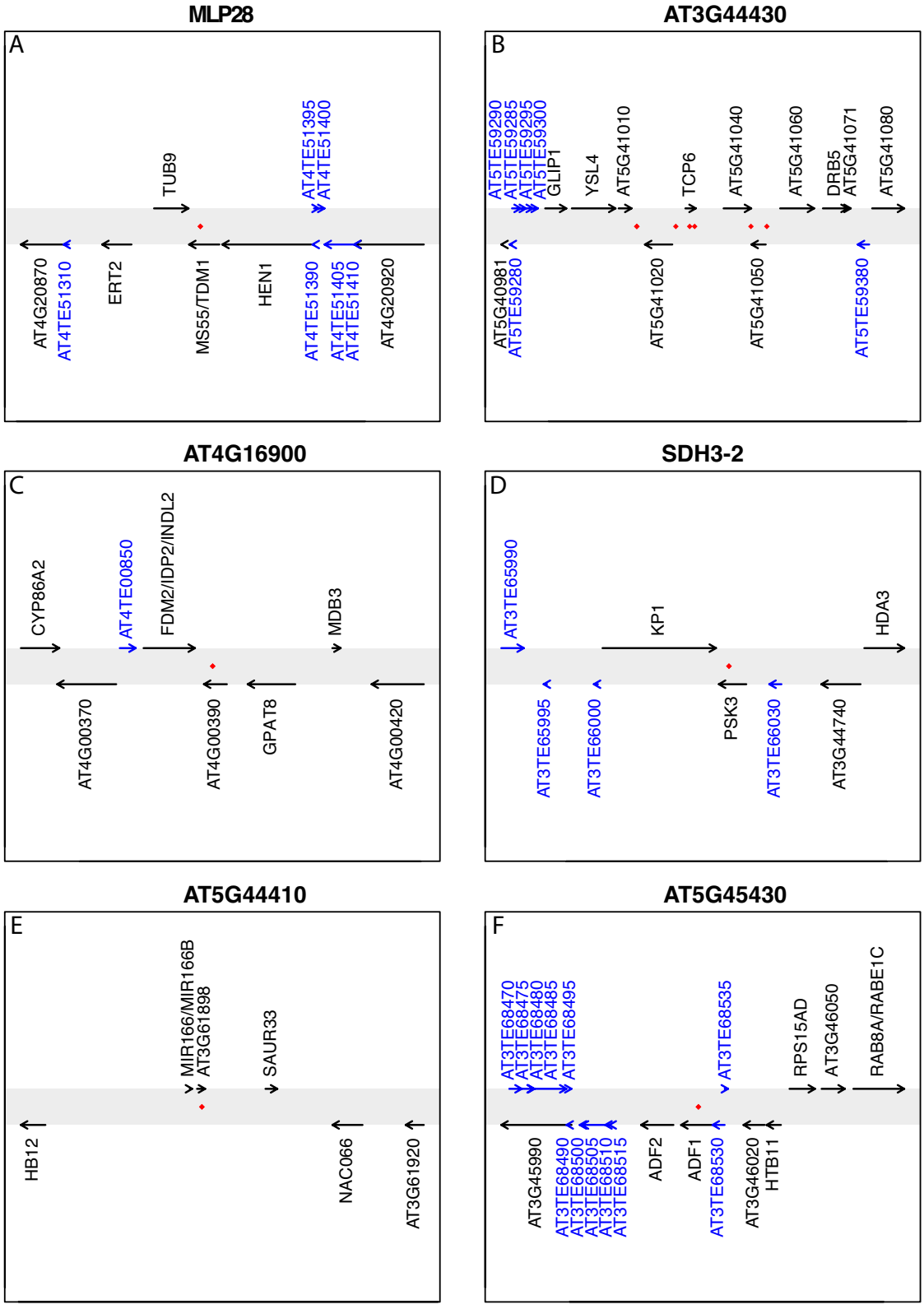


Figure 11: Schematic overview of associated SNP regions. Location of the SNPs (red diamonds) associated with treatment dependent expression of MLP28 (A), AT3G44430 (B), AT4G16900 (C), SDH3-2 (D), AT5G44410 (E) and AT5G45430 (F). Arrows indicate the direction of transcription, for genes in black and for transposable elements in blue.

The FAD-binding berberine family protein encoding AT5G44410 contains an FAD-binding and a berberine/berberine-like domain. The differential expression of AT5G44410 is associated with a SNP located in the unknown gene AT3G61898 (Figure 11 E).

Within 10kb of this SNP a number of other potential regulators are located: *miR166*, *SMALL AUXIN UPREGULATED RNA33* (*SAUR33*), *NAC066*, *HOMEOBOX 12* (*HB12*). The microRNA miR166 plays a role, together with miR165, in the regulation of shoot apical meristem, floral development (Jung and Park, 2006), leaf polarity (Ueno et al., 2007) and root cell fate (Carlsbecker et al., 2010). The SAUR protein family has often been linked to cell expansion in different plant organs and tissues (McClure and Guilfoyle, 1989; Gee et al., 1991; Chae et al., 2012; Spartz et al., 2012; Kong et al., 2013; Markakis et al., 2013). The function and role of SAUR33 is however not yet characterized. The transcription factor HB12 is known to induce cell expansion in leaves (Hur et al., 2015). Moreover, HB12 affects shoot and root growth in response to water deficit and does this together with HB7 as part of the ABA signaling pathway (Olsson et al., 2004). Another transcription factor within 10 kb of the associated SNP is NAC066, which is involved in secondary cell wall thickening. AT5G44410 is however not present among the known direct targets of NAC066 (Ruiqin et al., 2010).

The protein kinase AT5G45430 has not yet been functionally described. The differential expression of this kinase is associated with SNPs spread over four loci on chromosomes 1, 3, 4 and 5. Only on chromosome three the SNP was located in a gene body, that of ACTIN DEPOLYMERIZING FACTOR1 (ADF1; Figure 11F). Being involved in the turnover of filamentous actin. Within 10 kb of the associated SNP on chromosome three, the GTPase *RAB8A/RABE1C* is located. The RAB proteins are membrane organizers involved in the regulation of vesicle formation and transport (Zerial and McBride, 2001). Ectopic expression of *RAB8A/RABE1C* results in retarded shoot and root growth (Ahn et al., 2013). Other genes in the 10 kb region encode a galactose oxidase (AT3G46050), the ribosomal protein S15A D, the histone HTB11, an RNA-binding protein (AT3G46020), ADF2 and an actin-binding protein (AT3G45990). Theoretically each of the above listed genes may be a regulator of AT5G45430 transcription. None of the candidates does however play an obvious role in regulating gene expression. Therefore we cannot prioritize any of the candidates as the most plausible regulator.

Transposable elements as potential trait-specific regulators

Next to protein encoding genes, some of the associated SNPs were located in transposable elements (TEs). These elements may contain transcription factor binding domains or induce siRNA mediated mRNA degradation and thus influence gene expression of surrounding or further located genes (Rebollo et al., 2012). The differential expression upon drought associated with SNPs that were located in TEs for four of the eight genes (AT3G03700, AT3G44430, PLC8, AT5G45430; Figure 12 A-D). The detected TEs reside in different families of transposable elements but three of the five TEs are within the Helitron superfamily, which are characterized by a "rolling circle" transposition (Kapitonov and Jurka, 2007) and may rewire gene regulatory networks by supplying cis-regulatory elements (Ellison and Bachtrog, 2013).

Gene	Transposable Element	Family	Superfamily
AT3G03700	AT2TE00425	HELITRONY1B	RC/Helitron
AT3G44430	AT2TE28345	ATDNA12T3A	DNA
AT3G47290	AT1TE55145	HELITRONY1D	RC/Helitron
AT3G47290	AT1TE55100	ATREP4	RC/Helitron
AT5G45430	AT5TE40830	ATENSPM5	DNA/En-Spm

Table 5: Overview of genes for which the associated SNPs are located in transposable elements (TE) with indication of TE families and superfamilies.

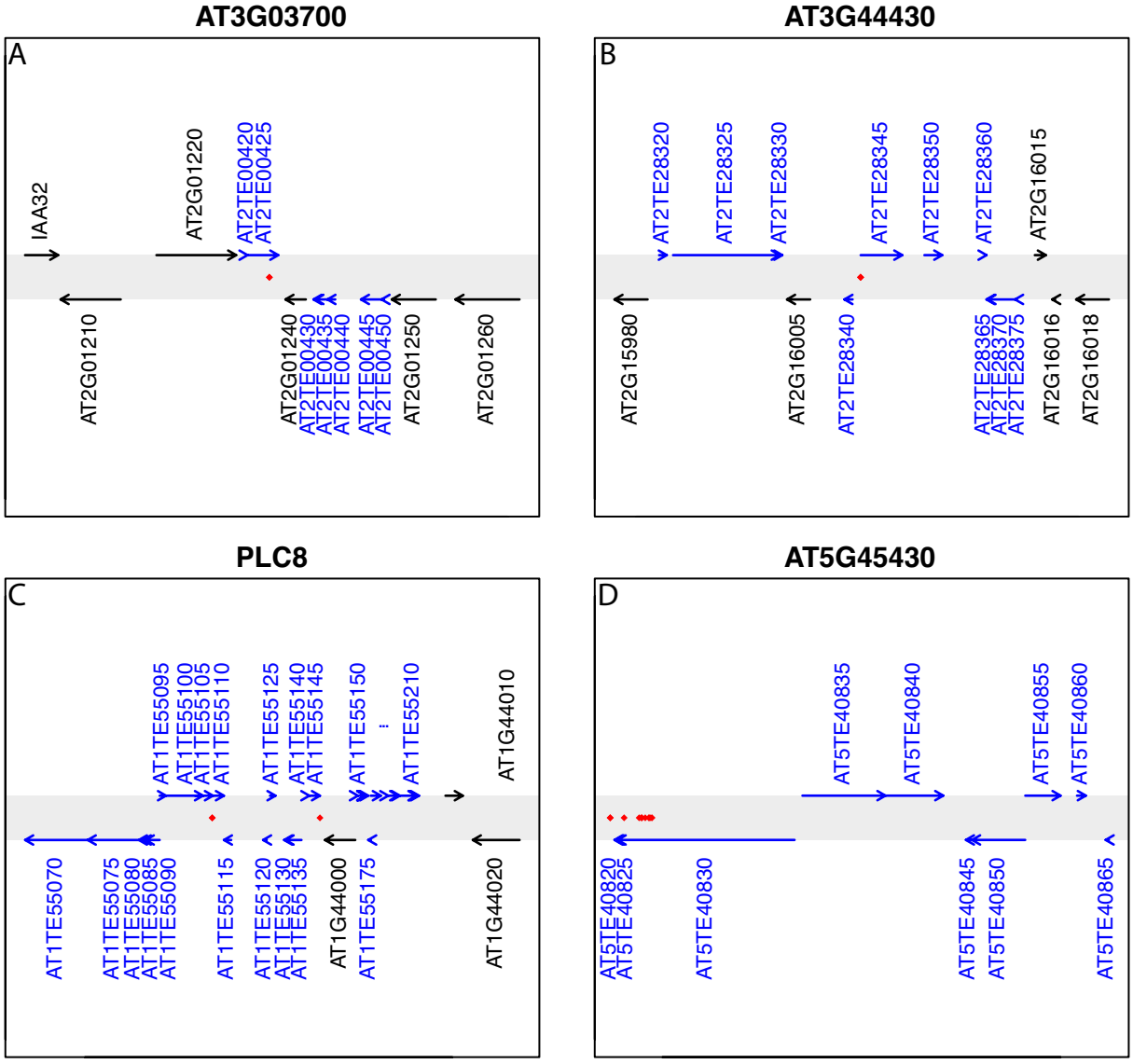


Figure 12: Schematic overview of associated SNP regions. Location of the SNPs (red diamonds) associated with treatment dependent expression of AT3G03700 (A), AT3G44430 (B), PLC8 (C) and AT5G45430 (D). Arrows indicate the direction of transcription, for genes in black and for transposable elements in blue.

Associated SNPs located in non-annotated regions

For two genes associated SNPs were located at loci with no annotated gene, TE, pseudogene or other genetic element. For AT3G44430, encoding a protein with unknown function, a SNP was detected upstream of AT2G42570 (Figure 13A), which encodes a trichome birefringence-like

protein. The birefringence-like genes contribute to the synthesis and deposition of cellulose in secondary cell wall formation (Bischoff et al., 2010). This protein is however unlikely to have a function in transcriptional regulation and the SNP might be in linkage with the true causal SNP located elsewhere. A good candidate within 10 kb of the SNP is *GENERAL REGULATORY FACTOR 9* (*GRF9*) gene that encodes a 14-3-3 protein with a regulatory signaling role in root growth and chloroplast development (Mayfield et al., 2012). Three SNPs on chromosome three associated to the differential expression of AT3G44430 (Figure 13B). Within 10 kb of these SNPs two genes encoding NAC transcription factor encoding genes are located (*NAC055* and *NAC056*), from which NAC055 is able to bind a drought-responsive element (Tran et al., 2004) and NAC056 is involved in the salt stress response (Balazadeh et al., 2010).

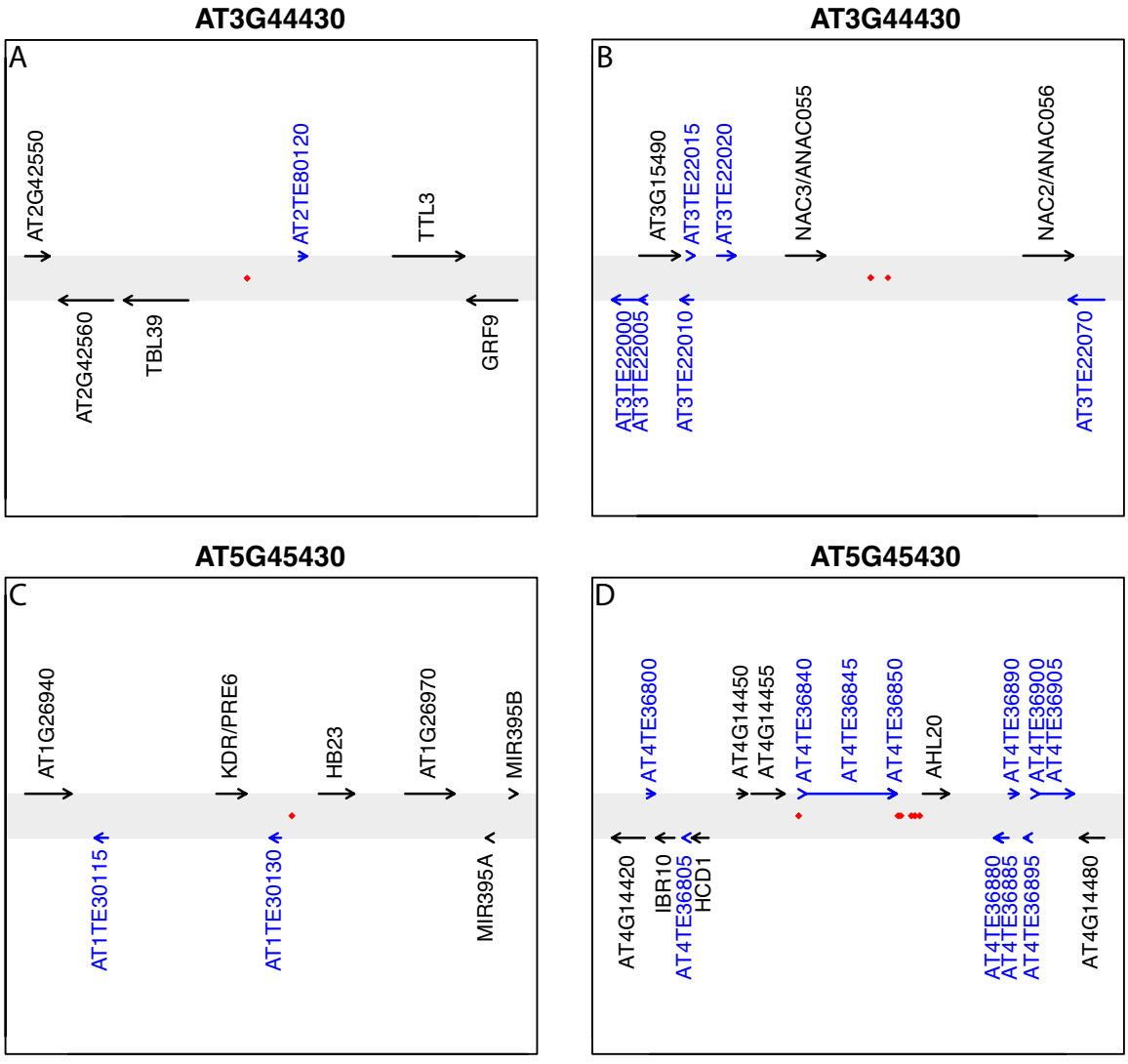


Figure 13: Schematic overview of associated SNP regions. Location of the SNPs (red diamonds) associated with treatment dependent expression of AT3G44430 with associated SNPs on chromosome 2 (A), AT3G44430 with associated SNPs on chromosome 3 (B) and AT5G45430 with associated SNPs on chromosome 1 (C). Arrows indicate the direction of transcription, for genes in black and for transposable elements in blue.

One of the SNPs that associated with the differential expression of the protein kinase AT5G45430 is located in the promoter region of the transcription factor HB23 (Figure 13C), recently described as one of the transcription factors that regulate osmotic stress tolerance by inducing proline accumulation (Kazama et al., 2014). Another group of associated SNPs located

on chromosome four and clustered in the promoter region of the gene encoding the AT-hook containing DNA-binding protein AHL20 (Figure 13D), which can repress expression of biotic stress genes (Lu et al., 2010). Moreover, members of the AHL family are involved in the regulation of plant growth (Zhao et al., 2013). Remarkably, one SNP is separated from the SNPs in the promoter region by three TEs. This SNP may be in linkage with the SNPs in the promoter.

Discussion

Phenotypic Variation among Accessions

Association analyses between phenotype and/or genome or transcriptome data requires sufficient phenotypic variation between the analyzed accessions. In the experimental set-up used in this study, the variation between the accessions was found to be significant for all phenotypes and for their responses to the mild drought stress. The difference between the extreme accessions was about four-fold for all phenotypes. The lack of a strong correlation between size reduction caused by mild drought stress and the size under control conditions for leaf 3 and for rosette area indicates that larger accessions are not per se more reduced under mild drought stress, nor do smaller accessions tolerate the stress better in terms of size reduction. This is in contrast to the weak but significant trade-off between plant size and the response to drought described earlier by analyzing 20 different accessions (Vile et al., 2011).

Previous reports have shown substantial natural variation in different ecophysiological traits in Arabidopsis such as flowering time and water use efficiency (Kenney et al., 2014); physiological effects of mild drought mature plants (Bouchabke et al., 2008); nitrate uptake and nitrate use efficiency (Chardon et al., 2010); heat and drought stress (Vile et al., 2011) and salt tolerance (Katori et al., 2010). Here we present a new set of growth related traits measured under control and mild drought conditions for 98 accessions.

Differences in cellular mechanisms determine final leaf size

Cell division and cell expansion are the two processes that determine leaf size. By analysing the 98 accessions, we found that mainly cell number correlates with the area of the leaf. The slight negative correlation between pavement cell area and cell number shows the existence of a clear trade-off between both parameters. Our data indicate the existence of three different strategies to produce larger leaves. In a first strategy, the mature leaf has an averaged number of cells with average sizes, leading to the formation of larger leaves. A second strategy is to produce larger leaves through larger, but less pavement cells. The lower cell numbers can be the result of either a limited number of cells that are recruited from the shoot apical meristem to the leaf primordium, a low cell proliferation rate or a shorter cell proliferation phase. On the other hand, larger cells can be the result of a more extensive cell expansion or a longer duration of the cell expansion phase. In the third strategy, a larger leaf area is reached mainly by investing in cell number whereas pavement cell areas are not explicitly larger than average. Here it is expected that more cells are recruited form the shoot apical meristem to form the initial primordium or that the cell division lasts longer or happens at a higher rate. Cell expansion rate on the other hand can be

lower or is less longer lasting. The higher pavement cell number can also be caused by an increase in meristemoid divisions. Meristemoids are the cells that give rise to stomatal guard cells and are spread out over the entire leaf (Peterson et al., 2010; Pillitteri and Dong, 2013). While the meristemoids divide asymmetrically during the production of guard cells, they also form pavement cells. As 48% of the pavement cells in the first leaf originate from meristemoids (Geisler et al., 2000; Bergmann and Sack, 2007), this mechanism can potentially have a large effect on final cell number and leaf size. The potential importance of the pavement cells formed through meristemoid divisions for the final morphology of the leaf is reflected in the surprising lack of correlation between the leaf size at the last day of full proliferation, when all cells are dividing, and the mature leaf size or pavement cell number. The division of the meristemoids is regulated by PEAPOD (PPD), a transcription factor that regulates a secondary cell cycle arrest front that involves the division of dispersed meristematic cells (White, 2006). Whereas the regulation of meristemoid division still requires further elucidation, much more information is available on the regulation of cell proliferation, cell expansion and the transition between these two pivotal leaf developmental phases.

For cell proliferation it is known that APC10, SAMBA and the GRF-INTERACTING (GIF) gene family can affect final leaf size by regulating, amongst other genes and proteins, the activity of important cell cycle regulators such as CYCA2;3 and CYCB1;1. (Lee et al., 2009; Eloy et al., 2011; Eloy et al., 2012). SAMBA targets CYCLIN A2 for APC/C-mediated proteolysis and leads to larger leaves and other organs when inactivated (Eloy et al., 2012), whereas the *gif1/2/3* triple mutant shows smaller leaves and shoot apical meristem (Lee et al., 2009). Also DELLA proteins and gibberelic acid (GA), which targets DELLA proteins for degradation, influence the cell proliferation rate as well as cell expansion (Huang et al., 1998; Coles et al., 1999; Achard et al., 2009; Gonzalez et al., 2010; Claeys et al., 2012; Dubois et al., 2013).

A large number of genes have been shown to have a role in the transition between cell proliferation and cell expansion. A key gene in the regulation of this transition is *miR319/JAW*. In the *Jaw-D* mutant, miR319 is overexpressed leading to larger and crinkled leaves (Palatnik et al., 2003; Ori et al., 2007). The leaf size increase is the result of a prolonged mitotic activity along the leaf's margins, leading to overgrown margins and hence larger leaves with more cells and a crinkled morhpology (Efroni et al., 2008). Regulation of the transition from cell divison to cell expansion through miR319/JAW is known to involve TEOSINTE BRANCED1/ CYCLOIDEA/PCF 4 (TCP4) (Palatnik et al., 2003), which in turn regulates miR396 and so affects *GROWTH REGULATING FACTOR* (GRF) transcripts (Rodriguez et al., 2010). Another transcriptional regulatory pathway in regulating transition involves the auxin-inducible AUXIN-REGULATED GENE INVOLVED IN ORGAN SIZE (ARGOS; Hu et al., 2003). *AINTEGUMENTA* (*ANT*) expression is suggested to be controlled by ARGOS (Hu et al., 2003) and could affect transition (Krizek, 1999; Mizukami and Fischer, 2000) through the regulation of CYCD3 expression (Mizukami and Fischer, 2000). Also a putative ubiquitin receptor, DA1, and a cytochrome P450, KLUH, can affect the duration of cell division (Anastasiou et al., 2007; Kazama et al., 2010; Stransfeld et al., 2010).

Just as for cell proliferation, also for cell expansion a number of genes are described that can alter the rate or duration of cell expansion. The best-known genes regulating cell expansion are the expansins. Their cell wall loosening activity facilitates cell expansion (Sampedro and Cosgrove,

2005). Overexpression of EXPA10 is known to produce larger leaves containing larger cells (Cho and Cosgrove, 2000). Also ARGOS-LIKE (ARL), TARGET OF RAPAMYCIN (TOR), SAUR19 and ZINC FINGER HOMEODOMAIN5 (ZHD5) result in larger leaves with larger cells when overexpressed (Hu et al., 2006; Deprost et al., 2007; Hong et al., 2011; Spartz et al., 2012). More details on the regulation of the different mechanisms involved in leaf growth can be found in (Gonzalez et al., 2012) and (Kalve et al., 2014).

The cellular analysis of the accessions indicates that there is a trade-off between cell division and cell expansion. The phenomenon by which cell expansion makes up for a reduced cell proliferation is referred to as compensation (Tsukaya, 2008) and indicates that there are organ-wide or even organism-wide mechanisms for the regulation of organ size, as shown for roots, petals and leaves (Tsukaya, 2002; Beemster et al., 2003; Ferjani et al., 2007). A set of mutants with clear compensation in cell expansion have been described (Ferjani et al., 2007) and allowed for defining three distinct classes of compensation. In class I, cells can enlarge post-mitotically by increasing their expansion rate; in class II the cell enlargement can result from prolonged cell expansion; and in class III an increased size of the dividing cells is observed (Ferjani et al., 2007). Although, the molecular pathways involved in class I & II remain unclear, there are strong indications for the involvement of KIP RELATED PROTEIN 2 (KRP2) in class III compensation (Ferjani et al., 2014).

However, the question remains why certain accessions invest more in cell division, while others rather in cell expansion. Cell size is well known to be correlated with nuclear genomic content, and artificially increasing ploidy levels results in larger cells (reviewed by Chevalier et al., 2014). Strikingly, the only known tetraploid accession, Wa-1 (Schmuths et al., 2004), is among the accessions with the largest cell sizes and has the largest leaf 3 area. Therefore it will be interesting to further investigate the ploidy level and endoreduplication index of the different accessions.

Compensation effects can be overcome by combining genetic alterations stimulating both cell division and cell expansion. Such combinations often result in synergistic interactions leading to leaves larger than the sum of the leaf sizes of both separate mutants (Vanhaeren et al., 2014). However, a combination of large and more cells has not been observed in the set of 98 natural variants. This argues for an organ-wide control on the size of the leaf, where a certain optimal leaf size is pursued. Such optimal leaf size is the complex integration of both intrinsic developmental signals and programs as well as a plethora of environmental factors (light, temperature, nutrients, water availability …), which make up for the leaf's economics (Wright et al., 2004), which ultimately defines the size of the leaf.

Drought affects both cell division and cell expansion

The mild drought protocol used in this study allows for the third leaf to be exposed to the mild drought stress during its entire development. The water availability was controlled for plants under both control and mild drought from 4 DAS onwards. The effect of the induced mild drought stress was already visible at the last day of full proliferation of the third leaf, with an average reduction of 21% over all accessions. In the course of the experiment, the water availability was further decreased at 11 DAS (detailed in Material and Methods). Since the development of the third leaf shifts from cell proliferation to cell expansion around 11 DAS, this

extra decrease in water availability could explain the stronger reduction in pavement cell area compared to pavement cell number, leading to a final leaf area reduction of 57% on average.

Reductions in cell expansion of plants exposed to drought stress have mainly been attributed to decreased turgor pressure (Maggio et al., 2006). Indeed, the Lockhart equation shows that the expansion of a cell depends on the extensibility of the cell, the internal turgor pressure and the turgor pressure needed for expansion (Lockhart, 1965; Green et al., 1971). In drought, plants will try to restore the turgor pressure by closing the stomata to avoid water loss and by producing osmolytes in order to increase the osmotic potential, the so-called osmotic adjustment (Zhang et al., 1999). Growing tissue will also loosen its cell walls in order to increase the extensibility of the cell wall, which also allows for expanding the cell at lower turgor pressures. Both expansins, xyloglucan endotransglycosylases and lowering the pH are playing a role in the drought induced cell wall loosening (Moore et al., 2008). A reduced turgor is also affecting cell division, since cells will only divide when a certain size is reached (Cleland, 1971; Taiz, 1984; Cosgrove, 1987; Beemster et al., 2005). Despite the possibility of plants to restore their turgor pressure, especially if the applied drought is rather mild as in our set-up, plants will still grow at a lower rate once the osmotic potential is restored (Matsuda and Riazi, 1981; Meyer and Boyer, 1981; Bressan et al., 1982; Bressan et al., 1990). This shows the existence of a regulatory mechanism that actively inhibits growth in drought conditions and is reflected in the adjusted but lower growth rate following an initial organ-wide growth arrest upon drought (Skirycz and Inzé, 2010). This active regulation of an adjusted growth rate under drought conditions is where growth and survival come in balance (Claeys and Inzé, 2013).

Adjustments in cell proliferation are regulated by affecting the CDK-cyclin modules at different levels. Three major mechanisms are controlling the CDK-cyclin complexes: cyclin degradation by ANAPHASE-PROMOTING COMPLEX/ CYCLOSOME (APC/C); inhibition or activation of CDK/cyclin activity through phosphorylation; interaction of CDK/cyclin complexes with the inhibitory CYCLIN-DEPENDENT KINASE INHIBITOR (CKI)/KIP-RELATED PROTEIN (KRP)-type proteins or SIAMESE (SIM)/SIAMESE-RELATED-type proteins. It is suggested that nearly all these components of the cell cycle machinery are affected by drought and that ethylene, abscisic acid (ABA) and GA are the main regulators. Regulation of the adjusted cell expansion rate is thought to involve auxin, ethylene and GA (reviewed by Claeys and Inzé, 2013).

The importance of reductions in cell division and cell expansion in determining the final leaf size under mild drought stress is shown by the correlations between the reductions in leaf three area and pavement cell number and area respectively, where a slightly greater significance was found for cell division. The alterations of cell division and expansion in response to mild drought stress have been described previously (Aguirrezabal et al., 2006; Pereyra-Irujo et al., 2008; Baerenfaller et al., 2012). Aguirrezabal et al. (2006) noticed that the extent to which cell division and expansion are affected depends on the drought severity. Cell division and expansion increased or decreased independently from each other in response to different drought scenarios in the An-1 and Cvi-0 accessions. Both processes were therefore suggested to be partly uncoupled (Aguirrezabal et al., 2006). Also in our study, in a large number of different genetic backgrounds, there is no correlation between the reduction in cell number and cell size. Although, under control conditions there was a trade-off between pavement cell number and area, which shows

that both processes are not regulated entirely independent. However, in response to mild drought stress, cell division and expansion seem to respond independently from one another in the different accessions. It is possible that the regulation of both processes is still coupled under mild drought stress but that the extent to which the two cellular processes respond is highly dependent on the accession.

Stomatal index in drought

Besides the reductions in pavement cell number and size, also the guard cells and thus stomata undergo changes under mild drought stress. Stomatal density is known to be influenced by environmental parameters such as CO_2 concentrations, humidity and light intensity (Bergmann and Sack, 2007). Also under increasing salt stress, different species decrease their stomatal density (Shabala, 2013). Adjustments of the stomatal number of developing leaves to changing CO_2 concentrations is influenced by a signal coming from mature leaves (Lake et al., 2001).

The 98 Arabidopsis accessions decreased their stomatal indices on average by 5%. Some of the accessions clearly reduced their stomatal index, while other rather increased it. Both increases and decreases of stomatal index upon drought stress have been reported for wheat (Quarrie and Jones, 1977), squash (Sakurai et al., 1986) and the grass *Leymus chinensis* (Xu and Zhou, 2008). In the latter study it was found that moderate drought stress causes an increase in stomatal index, whereas severe drought leads to a decrease. The exact reason why an increase or decrease of the stomatal index is beneficial in a certain environment still needs to be elucidated and will be the result of a complex interplay of stress responses of the cell division, cell expansion, photosynthesis, respiration and cuticle composition.

The lack of correlation between the stomatal index in control conditions and the reduction in leaf area shows that, although the amount of stomata are adjusted under mild drought, the number of stomata does not have a big impact on the tolerance against drought in terms of growth reduction. Nonetheless, stomata are important for plants to cope with drought. Mutant studies in severe drought conditions show that both stomatal density and stomatal aperture are contributing to drought tolerance (Yoo et al., 2010; Dong et al., 2014; Liu et al., 2014b). For the accessions tested here in mild drought it might be that the stomatal aperture contributes more to the drought tolerance of the accessions than the amount of stomata as such.

GWAS of mild drought responses of leaf growth-related phenotypes

With the recent availability of genotypic data of different accessions, GWAS is becoming a standard method in Arabidopsis to search for candidate genes that are involved in a certain trait. A first GWAS has been conducted on 107 different phenotypes using 96 accessions (Atwell et al., 2010). The associations were enriched for a-priori candidate genes, proving that many of the found associations are true. Nevertheless, also genes that were not known to be involved in the analyzed traits were found among the most significant associations. GWAS is hence a valuable tool to screen for new candidate genes that underlie a given trait. Since the paper by Atwell et al. (2010) many different traits have been subjected to GWAS, using populations of variable sizes. The initially used population of 96 accessions is generally considered as a minimal population size and recent studies are using populations counting between 179 up to more than 400 accessions

(Brachi et al., 2010; Li et al., 2010; Angelovici et al., 2013; Brachi et al., 2013a; Brachi et al., 2013b; Meijón et al., 2014; Slovak et al., 2014). In our set-up we opted to screen 98 accessions, which was at the moment of decision seen as a minimal but usable population size to perform GWAS (Atwell et al., 2010). Due to the lack of genotypic data of 7 accessions, the analysis finally made use of 91 accessions.

The performed GWAS specifically aimed at retrieving associations with the differential response to mild drought stress. This approach was successful in the sense that there were several loci that showed an association with a higher significance than the surrounding loci and this for different traits. However, none of the loci passed the significance threshold, except for one locus associated to the differential response upon mild drought stress of leaf 3 area at proliferation. This significant association is surprising since we found a lower heritability for this trait compared to the other phenotypes. Besides, the Manhattan plot (Fig. 4A) may indicate that the model is inflated and delivers more p-values of higher significance than expected. Therefore it cannot be excluded that the locus that significantly associates with leaf 3 area at proliferation is a false positive. Studies with a similar number of accessions in root architecture and root growth, traits expected to be of similar complexity as our leaf growth related traits, did however result in significant associations (Gifford et al., 2013; Rosas et al., 2013). The latter two studies did obtain very similar p-values as we obtained for the growth-related phenotypes. Because in GWAS the significance cut-off is corrected for multiple testing, the cut-off depends on the number of tests and hence the number of SNPs tested. Both studies on root traits tested approximately 241,000 SNPs obtained from a SNP chip, while in our study whole genome sequencing data (roughly 4,000,000 SNPs) was used. The multiple testing correction is therefore expected to be much more stringent in our study. Permutation tests would give a more correct estimate of the p-value cut-off at which a 5% false positive rate is expected.. Adding more accessions to the collection will also increase the statistical power of GWAS analyses to detect associations. More than 91 accessions are probably needed to detect significant associations with such a dense SNP dataset.

Nevertheless, the peaks that are visible in the Manhattan plots are linked to a number of genes that could have a sensible function in the specific trait. The locus that significantly associated to the response of the leaf 3 at proliferation to mild drought stress is linked to the genes *NLP7* and *TRE1*, both with a function in the drought stress response. The transcription factor NLP7 is mainly described for its role in nitrate sensing and metabolism, but is postulated to play a role in stomatal movement, leading to drought resistance in loss-of-function mutants (Castaings et al., 2009). However, if the role of NLP7 in the drought response is limited to the stomata, it is unlikely that NLP7 will play a role in proliferating leaves, as they lack stomata. Also in our transcriptome data, no evidence was found that NLP7 is differentially expressed in proliferating leaves. The neighboring drought related gene, *TRE1* encodes the Arabidopsis trehalase, the only enzyme known in this species to specifically hydrolyze trehalose into glucose (Muller et al., 2001; Lunn et al., 2006). Loss of endogenous trehalase by mutation results in increased trehalose levels (Van Houtte et al., 2013). Therefore, natural variation in TRE1 functionality or expression levels are likely resulting in varying levels of trehalose, which is a known osmoprotectant (Elbein et al., 2003). Besides, TRE1 was shown to regulate stomatal index (Vandesteene et al., 2012; Van Houtte et al., 2013). Hence, TRE1 is a good candidate for explaining the differential response of the proliferating leaf to mild drought stress.

Although the other growth-related phenotypes did not show statistical significant associations, some peaks in the Manhattan plots were visually obvious. For instance, leaf 3 area at maturity associated clearly with a locus near *miR171c*. The miR171c-targeted scarecrow-like proteins (SCL6/22/27) negatively regulate chlorophyll biosynthesis via an unknown mechanism. SCL proteins inhibit the expression of the key gene encoding protochlorophyllide oxidoreductase (POR) in light-grown plants and the miR171-SCL module is critical for mediating GA-DELLA signaling in the coordinate regulation of chlorophyll biosynthesis and leaf growth in light. Hence, miR171c may also be involved in regulating leaf growth upon mild drought stress.

Genes that are involved in the differential response of the screened phenotypes may show differential expression upon the mild drought treatment. The genes that associated with the different phenotypes (as indicated in Figure 4 and Supplementary Figure 3) were checked for differential expression upon mild drought stress (fold change > 2). Genes that are associated with the mild drought response of rosette area and also are differentially expressed by mild drought stress encode the following proteins: two receptor-like proteins (RLP39 and RLP40), the xyloglucan synthesizing CSLC4 and the zinc-finger transcription factor SAP12. To our knowledge the two RLPs have not yet been functionally described. CSLC proteins, are thought to synthesize the β-(1→4)-linked glucan backbone of xyloglucan, an abundant polysaccharide in the primary walls of many plants (Davis et al., 2010). Under drought stress, growing tissue will be kept extensible so that the cells can expand with less turgor pressure (Wu and Cosgrove, 2000; Clauw et al., 2015). Xyloglucan is the main hemicellulose in plants and by linking cellulose microfibrils, xyloglucan is important for the strength of the cell wall. By cutting and grafting the xyloglucans, the cell wall modifying xyloglucan endotransglycosylases/hydrolases (XTHs) modify the extensibility of the cell wall and hence the ability to grow. It is likely that the associated SNPs in the *CSLC4* gene have an effect on the functionality of xyloglucan or the amount that is synthesized, which both influence the extensibility of the cell wall and hence the ability to grow under mild drought. The *SAP12* gene is strongly induced under cold and salt stress in a time-dependent manner (Ströher et al., 2009) and may also have a function in the mild drought response of rosette growth. However its functional role in stress responses needs further investigation.

Also *UFO*, associated with pavement cell number, was more than two fold differentially expressed upon mild drought in proliferating leaves. *UFO* encodes an F-box protein and UFO interacts with the transcription factor LEAFY (LFY) to regulate cell proliferation of shoot apical and inflorescence meristems (Risseeuw et al., 2013). UFO acts a co-activator of LEAFY and is required for proper activation of APETALA3 in the floral meristem during the specification of stamens and petals (Risseeuw et al., 2013). Although the role of UFO in leaf growth is not yet established variations in the amount of cells that are recruited from the shoot apical meristem to form the leaf primordium may affect the final leaf size and cell number (Gonzalez et al., 2012), a process which is plausibly affected by mild drought.

A thioredoxin encoding gene (AT1G60740) was found to be differential expressed and associated with the response of stomatal index to mild drought. Thioredoxins are generally known as important scavengers of reactive oxygen species, which are often produced upon abiotic stress (Vieira Dos Santos and Rey, 2006). The thioredoxin AT1G60740 is thus likely involved in the mild drought response but without an obvious stomatal-specific function. Also

associated with stomatal index was *EXORDIUM-LIKE1* (*EXL1*), a gene involved in the regulation of growth when the carbon availability is reduced (Schröder et al., 2011; Schröder et al., 2012). Variation in stomatal index in response to mild drought may indeed cause differences in the efficiency of CO_2 influx. Variations in EXL1 function or expression may form an adaptation for these differences in carbon availability, which may explain the detected association.

In conclusion, the GWAS delivered a number of candidate genes for the differential response to mild drought of the different traits. However, significance was reached for only one trait. The complexity of the growth-related traits and the high number of SNPs tested require more accessions to draw statistical sound conclusions. Nevertheless, a number of potential interesting candidates are suggested for which further functional characterization is required to validate the role of the detected genes in the specific traits they associate with.

Variation in the transcriptome of the proliferating leaf in response to mild drought stress

Measuring transcriptional differences between Arabidopsis accessions has indicated substantial variation in gene expression in different conditions (van Leeuwen et al., 2007; Delker et al., 2010; Marais et al., 2012). Also in our specific experimental set-up considerable expression variation was observed between the different accessions. Changing gene expression may be important for the adaptation to mild drought stress in the different accessions. For some traits the expression variation of the regulatory network can be more important than functional polymorphisms, as was shown for the regulatory network of auxin responses (Delker et al., 2010). Adaptation by gene expression variation may be especially important in quantitative traits where the trait is *in se* not differing in functionality between accessions, but rather in characteristics of the function such as the rate, size or duration. In a simplistic example, an increase in growth can be caused by higher expression levels of a growth-regulating gene, which causes a higher growth rate without a change in the growth regulatory function.

Although a large variability in differential expression was detected between the accessions, a large part of the varying genes clustered together for their response to mild drought over the different accessions. Co-expression and by extension co-differential expression has been shown to point towards transcription factor binding site similarity and thus co-regulation of the clustered genes (Allocco et al., 2004) and suggests functional relatedness of the clustered genes. For the small set of clusters that we checked in more detail, the genes were indeed functionally related to some extent and are thus likely also co-regulated. However, co-expression as such does not deliver the potential regulators. In order to do so the eGWAS analysis was conducted, which is discussed below.

Despite the large number of genes that do vary in expression between accessions, there is a set of genes that are responding similarly to the mild drought stress in 90% of the accessions. Of the eleven genes in this core set, ten were previously described to show a common expression upon mild drought stress in a smaller subset of six accessions (Clauw et al., 2015). A similar core transcription response of 11 genes has been detected for iron deficiency (Stein and Waters, 2011). Although differences in functionality of these core responsive genes cannot be excluded, the

conservation of their transcriptional response can indicate that they are important players in the mild drought response of young developing leaves.

Stress predictors

To prioritize genes that show differences in expression between control and mild drought stress, a modeling approach was used. We built a classification model that was able to distinguish control from mild drought stress samples, based upong the expression values of a set of 283 genes. In other words, the expression values of this group of 283 genes together are characteristic for the mild drought treatment. Therefore, these 283 genes are referred to as stress predictors. The genes were enriched for drought stress related genes, from which 79 were previously described to be involved in the mild drought response of young developing leaves (Clauw et al., 2015). Among the stress predictors were 28 genes involved in ABA signaling and responses, which are commonly known to be involved in drought responses. In addition to genes associated with the phytohormone ABA, there was also a clear indication of the involvement of gibberellic acid (GA). More specifically *GA INSENSITIVE DWARF 1B* (*GID1B*) encoding a GA receptor and two genes encoding GA biosynthesis genes (*GA20OX1* and *GA20OX2*) were among the stress predictors. The GA 20-oxidase enzymes, encoded by *GA20OX1* and *GA20OX2*, catalyze the sequential oxidation of GA_{53} to GA_{20}, which is subsequently hydroxylated to the bioactive GA_1 by GA 3ß-hydroxylase. The induction of the expression of genes encoding GA 20-oxidases leads to increased GA levels (Hisamatsu, 2005). Bioactive GA can then bind the GID1 receptors leading to a conformational change that allows for complex formation with a Skp-Cullin-F-box (SCF) complex. This GA-GID1-SCF complex is responsible for the degradation of DELLA transcription factors, which will activate the GA transcriptional response. DELLA proteins are of pivotal importance to regulate plant growth in a tissue and organ dependent manner (Claeys et al., 2014).

Another interesting set of stress predictors was a group of three bHLHs (bHLH38, bHLH39, bHLH100) that were co-differentially expressed with a gene that encodes a protein of unknown function (AT2G14247). The CAST clustering, which was also used for co-differential expression, found also bHLH101 to cluster with bHLH39, bHLH39 and bHLH100. The four bHLHs are involved in the regulation of the transition from cell proliferation to differentiation during leaf development. The expression levels of the bHLHs are regulated by the repressor TCP20 and the activator, SUPPRESSOR OF ACAULIS 51 (SAC51; Andriankaja et al., 2014). Through the repressive effect of TCP on genes involved in differentiation, TCP20 appears to stimulate cell proliferation (Li et al., 2005). Taken together, the data suggest that bHLH39, bHLH39, bHLH100and bHLH101 play a key role in stimulating upon mild drought stress the transition of proliferating cells to expanding cells. This is in agreement with the expression changes of genes involved in cell wall modification seen in six accessions, where genes encoding cell wall loosening proteins are up-regulated under mild drought in preparation for cell expansion (Clauw et al., 2015).

Finally, the modeling approach succeeded in retrieving a valuable set of genes based on the natural variation in the transcriptional response upon mild drought stress. For some of the genes a function in the response could be suggested based on what is known in literature. The genes

without an obvious function in drought or growth-regulation are interesting candidates for further investigation.

eGWAS

The substantial gene expression variation between the 89 accessions allowed for associating genetic polymorphisms to the expression differences in an approach similar to eQTL mapping but using a collection of natural accessions instead of recombinant inbred lines. The collection of natural accessions contains more genetic diversity and gives a higher mapping accuracy due to the higher number of recombinations in natural accessions in comparison to a standard biparental RIL population. An extra complication in association studies using natural populations is the population structure, which was tackled by adding the kinship matrix as random factor in the mixed model. With our main interest being the differential response upon mild drought stress, a multi-trait mixed model (Korte et al., 2012) was used.

This approach proved successful at retrieving loci that associated with treatment independent gene expression (common test) and differential expression upon mild drought stress (trait-specific test). However, a substantial number of genes for which the eGWAS was performed, showed inflation of the model. Inflated models typically result in a large number of significant associations that are biologically not meaningful. Therefore we chose to remove these genes from further analyses. The selection was based on the genomic inflation factor (λ) and the D'Agostino test. Because the D'Agostino tests for a potential cause of inflation (non-normal distribution of expression values) but not inflation as such, it is probably overly stringent. Therefore we expect that the D'Agostino selection also excluded genes for which the model was not inflated from the analysis. Nevertheless, the selection based on these two criteria resulted in a small set of genes that showed associations with their expression values and did not suffer from model inflation. The high stringency of the selection also makes it highly plausible that the selected associations are true positives.

The number of genes with significant associations differed greatly between the common test on treatment independent expression (95 genes) and the trait specific test on differential expression upon mild drought stress (8 genes). The common test used expression values of both treatments as repeats and possessed therefore more statistical power than the trait specific test where the number of expression values per haplotype halves. Among the associated loci in the common test, the great majority was located in the proximity of the regulated gene. These SNPs are mainly thought to be *cis* regulators but the possibility of nearby *trans* regulators cannot be excluded. Typically, *cis* located regulators have a greater phenotypic effect than *trans* regulators (Meiklejohn et al., 2003; Yvert et al., 2003; Raser and O'Shea, 2004; Wayne et al., 2004; Brem and Kruglyak, 2005; Hughes et al., 2006; Keurentjes et al., 2007; West et al., 2007) and thus require less statistical power for significant association. Therefore, the combination of higher statistical power with the high number of *cis* regulators in the common test may explain the higher number of genes for which associations are found in comparison to the trait specific test.

Due to the lower statistical power in the trait-specific test, fewer associations are expected and especially fewer associations in *trans* located loci. However, more *trans* loci were detected in the trait-specific test, and *cis* loci were even completely absent. Therefore we postulate that

polymorphisms in *trans* regulatory loci mainly shape the genotype specific expression response to specific environmental perturbations. Moreover, the *trans* loci detected in the trait-specific test had a larger phenotypic effect than the *trans* loci in the common test. Further functional characterizations of the *cis* and *trans* loci will need to elucidate why adaptations to environmental perturbations are affecting *trans* rather than *cis* elements.

Functional diversity among the potential regulators

The potential regulatory loci that were detected in the eGWAS analysis were functionally diverse. Remarkably, the associated *cis* loci were only detected as potential regulators of treatment-indepedent gene expression. As expected, the greatest density of *cis* loci was detected in a region up to 2kb upstream of the transcribed region (as defined in TAIR10; www.arabidopsis.org). This strongly indicates the importance of promoter elements such as transcription factor binding sites and enhancers or silencers in explaining the observed differential gene expression. The lowest density of associated loci was located downstream of the transcribed region, with an intemediate level of associations in the gene body itself. The SNPs in the gene body might affect gene regulation by interfering with, for example, alternative splicing or alterations of the stability of an autoregulated protein. SNPs were also detected further than 2kb from the gene, these can be located in more distant enhancer regions. Alternatively, it cannot be excluded that proximate *trans* loci are playing a role in the expression regulation. Functionally related genes can co-localize and also the regulators may thus be present in the proximity of such a cluster of genes (Schweizer and Stein, 2011).

The *trans* loci were mainly associated with differential expression upon mild drought, but also *trans* loci associated with the treatment independent expression of three genes. The associated SNPs were located in or close to different functional types of genetic elements such as protein encoding genes, siRNA encoding genes, transposable elements, and promoter regions.

For the unknown gene AT3G44430 and SDH3-2, encoding a membrane anchor for the succinate:ubiquinone oxidoreductase complex, the SNPs that associated with their differential expression were located in genes with a plausible regulatory function. For AT3G44430 the plausible *trans* regulator is *TCP6*, encoding a transcription factor of the TCP family that is involved in plant development and specifically in cell proliferation (Martín-Trillo and Cubas, 2010). Differential expression of *SDH3-2* was associated with *PSK3* encoding one of the five preproteins the peptide hormone phytosulphokine (PSK), which is involved in the regulation of cell longevity and cell proliferation (Matsubayashi et al., 2006). The PSK preproteins PSK2, PSK3, PSK4 and PSK5 (Yang et al., 2001; Lorbiecke and Sauter, 2002) are widely expressed in cotelydons and leaves, and are generally thought to be transcriptionally regulated by developmental and environmental signals (Sauter, 2015). ERF115 is a transcription factor that regulates PSK in controlling cell division in the root quiescent center (Heyman et al., 2013). The role of PSKs in cell division is the possible explanation why a SNP in *PSK3* associated with the expression of the gene encoding the mitochondrial protein SDH3-2. As a membrane anchor for the succinate:ubiquinone oxidoreductase complex, SDH3-2 fulfills a crucial structural function in the mitochondria. During cell division mitochondria undergo substantial morphological changes and fuse to form a 'cage' surrounding the nucleus. These large mitochondria produce the energy needed for cell division and during this 'cage' phase mitochondrial contents are redistributed.

Eventually the large structure divides and smaller mitochondria are divided over the newly forming cells (Kianian and Kianian, 2014). During such radical morphological changes it is very likely that a structural protein such as SDH3-2 is playing an important role. The leaf samples harvested in this study for transcript profiling consists of a large proportion of dividing cells and mild drought stress has a profound effect on cell proliferation. It is therefore not surprising that SDH3-2 is differentially expressed in response to mild drought stress. PSK may thus be a regulator of cell division and thus affect the structural dynamics of mitochondria such as changes in SDH3-2 expression, upon mild drought stress. The regulation of mitochondrial processes by regulators of cell division has previously been shown for Cdk1/CycB in mammalian cells (Taguchi et al., 2007).

As discussed earlier, so-called linkage disequilibrium might result in the observation that a SNP in close proximity to the actual causal SNP associate with the phenotype. Linkage disequilibrium decays within 10 kb in Arabidopsis (Kim et al., 2007). For AT4G16900, *MLP28* and AT5G44410 a sensible regulatory gene was detected within 10kb of the associated SNPs, rather than at the location of the associated SNP itself. The SNP associated with differential expression of the disease resistance gene AT4G16900 is located in a DNA-binding storekeeper protein with a known function as transcriptional regulator. However, the expression of the cluster of neighboring R genes of which AT4G16900 is part, is suggested to be epigenetically regulated (Stokes et al., 2002). Therefore, one or both of the two genes that have a function in DNA methylation (*MBD3*, *FDM2/IDP2/INDL2*) that map in the close proximity of the associated SNP are more plausible candidate regulators of the expression of AT4G16900 than the DNA-binding storekeeper protein. For *MLP28*, the associated SNP located in *MS5/TDM1*, which is involved in meiotic cell division. However, in the 10kb region the HEN1 gene is located. HEN1 encodes a general miRNA stabilizer and is the rate limiting factor of miR157d and miR319/JAW accumulation (Tsai et al., 2014). JAW infuences leaf development through regulation of TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR1 (TCP) transcription factors (Palatnik et al., 2003). TCPs, in turn, affect leaf development by regulating cell proliferation (Martín-Trillo and Cubas, 2010). Because of this link with regulation of cell proliferation, an active process in the sampled leaves, HEN1 is a more plausible candidate regulator than MS5/TDM1. Also for the FAD-binding berberine family protein encoding AT5G44410, the associated SNP locates in the gene of unknown function, AT3G61898, which is not the most plausible regulator. HB12, located within 10kb of the associated SNP is more likely a candidate for the regulation of AT5G44410. Not only is HB12 a transcription factor, it also regulates shoot and root growth in drought, interaction with HB7 (Olsson et al., 2004). Furthermore, HB7 is differentially expressed upon mild drought stress in young developing leaves (Clauw et al., 2015),is one of the stress predictors and is one of the eleven signature genes that show a similar expression response to mild drought stress in more than 80 accessions.

For four genes (AT3G03700, AT3G44430, PLC8, AT5G45430), the associated SNPs were located in transposable elements (TE). TEs often possess potential regulatory elements and as mobile genetic elements they have an important role in evolutionary adaptation. TEs are known to cause differences in gene expression between *Arabidopsis thaliana* and *Arabidopsis lyrata* (Hollister et al., 2011). TEs can influence gene expression in different ways, they often contain transcription factor binding elements and can function as promoters for neighboring genes. Since

the TEs identified in this study are located in *trans*, this implies that the TEs regulate the expression of a neighboring gene that in turn would be the actual *trans* regulator. However the associated TEs were not located in the promoter region of a gene with an obvious gene regulatory function. TEs may also encode for non-coding RNAs such as miRNAs and long non-codingRNAs (lncRNAs) which are known to have gene regulatory functions (Hadjiargyrou and Delihas, 2013). Different miRNAs and lncRNAs are involved in drought responses (Song et al., 2013; Di et al., 2014) and can influence gene expression and alternative splicing (Bardou et al., 2014). The location of these TEs is however based on the Col-0 reference annotation. Seen the mobility of TEs, the question is to what extent these TEs are present in the different accessions and how their function as transcriptional regulators can be interpreted.

The different potential regulators cover different biological functions. The gene functions of *trans* regulators are mainly expected to be transcription factors. From the eGWAS a number of transcription factors are indeed candidate regulators such as TCP6 and HB12. Beside the transcription factors, also a peptide hormone preprotein encoding gene (*PSK3*), genes involved in DNA methylation and TEs are suggested as potential transcriptional regulators. As a result of the low number of detected *trans* loci it is hard to make a general statement on the functionality of *trans* regulators. A previous genome-wide study in yeast could however show that only 17% of the *trans* regulators were transcription factors (Yvert et al., 2003). A possible explanation is that not one but multiple transcription factors are regulating the differential expression. Hence, the combination of multiple SNPs will correlate with the expression but not each SNP individually. This could be overcome by using a multi-locus mixed model (Segura et al., 2012).

Improvements and future directions

Potential regulators of differential expression were found for only eight genes of the in total 3393 genes on which the eGWAS was run. As discussed the reason for this low number of genes is that a very stringent selection was conducted to avoid the selection of genes for which the model inflated. This selection resulted in only 371 genes that were used in subsequent analyses and therefore the prevention of model inflation is a first measure to increase the amount of detected regulatory interactions. As the inflation could be the result of non-normal distributed gene expression values across the different haplotypes, increasing the number of haplotypes and thus the number of accessions is thus very likely to improve the performance of eGWAS and hence the number of regulatory interactions that will be detected.

Another observation is the larger number of regulatory interactions retrieved in the common test compared to the trait-specific test. This will be partly due to increased power for the common test. However, also the gene selection may have a role. The genes for the eGWAS were selected based on high heritability in control or mild drought conditions. In other words, these genes had a strong genetic component that explained their expression variation in either of both treatments, increasing the chances to detect the genetic component responsible for the expression differences between haplotypes. This does however not implicate a strong genetic component for the differential expression of these genes. Moreover there is no indication that these genes are involved in the mild drought response. Yet, the choice for selecting on heritabilities is defendable as this was a first attempt to perform such analysis and a high heritability is a necessity to detect significant associations. To improve the eGWAS and especially the trait-specific test, it may be

worthwhile to select genes that show differential expression and/or show a high heritability for their differential expression. Also the stringent selection against inflated models may be improved. As discussed above, this selection most probably excluded a number of informative associations. The final selection of 371 genes is therefore most probably only the tip of the iceberg. Many more regulatory interactions are hidden in the data, and it is a matter of finding the right method to unveil them.

A possibility is to loosen the selection against inflated models and add *a priori* biological knowledge to the eGWAS results in order to build a gene expression network containing new candidate regulatory interactions. Information on connectivity between different nodes in the network can come from different information sources. Literature can suggest interactions between different nodes and give information on the types of interaction. Co-expression analysis may suggest new nodes in the network as being co-regulated. Large transcriptome and genome datasets are available or are being produced nowadays. The statistical methodology to perform eQTL mapping and eGWAS have been developed and text mining tools such as EVEX (Van Landeghem et al., 2012) allow for high-throughput screening in literature for known regulatory interactions between genes. Each of the methodologies has its limitations. Literature is never complete and high-throughput text mining has an efficiency of 60% (Van Landeghem et al, 2013), eGWAS does not prove causality and results may contain false positives, eQTL mapping is limited in diversity and may deliver less regulatory interactions, co-expression analysis hints for co-regulation but does not give the regulator. However, the different methodologies are very complementary and seen the recent advances it is only a matter of time before someone will integrate them.

The approach suggested above mainly focuses on the *trans* regulations, however, *cis* regulator information may add extra proof for regulatory networks. The regulatory SNPs in *cis* may point towards important transcription factor binding sites or to specific enhancer elements. Identifying the underlying transcription factor binding sites of the associated *cis* loci will be very informative. Knowledge on the transcription factor binding sites can suggest a transcription factor as regulator. Co-expressed genes with the same transcription factor binding element associating to their expression may then potentially be regulated by the same transcription factor. However, variation in *cis* regulation may not always be as simple as a difference in transcription factor binding site. This is recently exemplified by a study on the natural variation of FLOWERING LOCUS T (FLT) expression. FLT is regulated by the interaction of an enhancer region with a transcription factor binding site region. The efficiency of this interaction defines the extent to which FLT is expressed. The interaction between the enhancer and transcription factor binding site is mediated by a region that is located in-between the two regulatory elements. To interact, the two regulatory elements need to bind this intermittent region, upon which the expression is initiated. Sequences of the enhancer, transcription factor binding site and the intermittent region are conserved between accessions that differ in FLT expression, but the distance between the intermittent region and the regulatory elements differed and defined the efficiency of the interaction and hence FLT expression (Liu et al., 2014a).

In conclusion, we successfully performed an association study of gene expression in a collection of natural accessions. Treatment-dependent and -independent expression clearly differed in the type of regulatory loci, where the treatment independent contained mainly *cis* located loci in

contrast to only *trans* located loci for the treatment dependent expression. The majority of the potential *trans* regulators of treatment dependent expression could be linked to growth regulation. Due to the expected increase in availability of this type of datasets, similar analyses will most likely become a standard approach to unravel transcriptional regulatory interactions. The most logic step further is to use this data in constructing gene regulatory networks, integrating different data sources to build networks that can give us more insight in the transcriptional regulation in specific tissues and environments.

Material and methods

Plant growth conditions and experimental setup

A collection of 98 accessions was grown in a growth chamber under controlled environmental conditions (21°C, 55% relative humidity, 16-h day/8-h night, and 110– 120 μmol m^{-2} s^{-1} light intensity). All accessions were bulked simultaneously. The last day of full proliferation varied between 8-10 days after stratification (DAS) depending on the accession and was determined microscopically as the last day where no expanding cells were observed at the leaf tip. Plants were subjected to a controlled drought treatment that started at 4 DAS (described in Clauw et al., 2015). The soil water content was controlled by using the automated phenotyping platform WIWAM (Skirycz et al., 2011). The control treated pots contained 2.2 g water g-1 dry soil, drought treated pots initially contained 1.2 g water g^{-1} dry soil and decreased to 0.7 g water g^{-1} dry soil after 10 DAS. Experiments were conducted in batches of 16 accessions. Two reference accessions (Col-0, Oy-0) were added to each batch to correct for batch effect. Experiments to obtain the growth-related phenotyping data were conducted in duplicate. For transcriptome analysis only one repeat was done. Images of the rosette of each plant were taken daily until 22 DAS and analyzed for the projected rosette area (PRA) (Skirycz et al., 2011). Size measurements of the third leaf were done at the transition from cell proliferation to cell expansion (10–11 DAS) and at maturity (23 DAS). For practical reasons, the mature third leaf was harvested 1 day later than the last PRA measurement at 22 DAS. To this end, the leaves were cut from the rosette, cleared in ethanol, and transferred to lactic acid before mounting on microscope slides. Measurements based on microscope images were done using ImageJ (http://imagej.nih.gov/ij/), and analysis of drawings made from the abaxial epidermis allowed for quantification of the cell area, cell number, and stomatal index (Andriankaja et al., 2012). All correlations with pavement cell number were conducted with exclusion of the outlier accession Ler-1.

Normalization of the phenotyping data

Because accessions were grown in separate batches, the phenotyping data was normalized to correct for batch effects. To this end a mixed model was used with genotype, treatment and their interaction as fixed factors. The batch effect was added to the model as a random factor:

$$y \sim batch + genotype + treatment + genotype * treatment + \varepsilon$$

The least square means over the two replicates for each genotype and treatments were then used in the further analysis. The normalization was performed with the lmer function in the R package "lme4" (v. 1.1-6), the least square means were obtained with the lsmeans function in the R package "lsmeans" (v. 2.10). All analysis were performed using R version 3.0.1.

Transcriptome analysis

Sampling

To ensure sufficient material for transcriptome analysis, 60 seedlings were grown per accession per treatment. Plants were harvested at the last day of full proliferation as was determined for each accessions through microscopic analysis. The time of harvest therefore ranged from 8 to 10 DAS (Days After Stratification) depending on the accession. Plants were flash-frozen in liquid nitrogen upon harvest. To prevent RNA degradation, RNA*later*®-ICE (Ambion®) cooled at -70°C, was added to the samples and was allowed to penetrate the tissue at -20°C for 5 days. The third leaf was collected by microdissection under a microscope. Samples were microdissected in a Petri dish on dry ice to keep the samples below room temperature. Dissected leaves were ground with a Retsch machine (Retsch) and 3-mm metal balls.

RNA extraction

RNA was extracted with TriZol (Invitrogen) according to the manufacturer's instructions. RNA samples were subjected to DNA digestion with RNase-free DNase I Kit (Qiagen), and impurities were removed with the RNeasy Mini Kit (Qiagen).

RNA-sequencing

Library preparation was done using the TruSeq RNA Sample Preparation Kit v2 (Illumina). In brief, polyA containing mRNA molecules are reverse transcribed, double-stranded cDNA is generated and adapters are ligated. After quality control using 2100 Bioanalyzer (Agilent), clusters are generated through amplification using the TruSeq PE Cluster Kit v3-cBot-HS kit (Illumina) followed by sequencing on Illumina HiSeq2000 with the TruSeq SBS Kit v3-HS (Illumina). Sequencing was performed in Paired-End mode with a read length of 50 nt. The quality of the raw data was verified with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, version 0.9.1). Next, quality filtering was performed using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/, version 0.0.13): reads where globally filtered in which for at least 75% of the reads the quality exceeds Q10. Adapters were trimmed using cutadapt (adapter sequence 'GATCGGAAGAGCACACGTCTGAACTCCAGTCAC' with at least 10 nt overlap; Martin, 2011). Next, 3' trimming was performed to remove bases with a quality below Q20, ensuring a minimum length of 35 nt remaining. Re-pairing was performed using a custom perl script. Reads were subsequently mapped to the Arabidopsis reference genome (TAIR10) using GSNAP (Wu and Nacu, 2010), version 2013-02-05) allowing maximally 2 mismatches. The concordantly paired reads that uniquely map to the genome where used for quantification on the gene level with htseq-count from the HTSeq.py python package (Anders et al., 2014).

Expression data normalization

RNA-Seq data were normalized for library size using DESeq2 v.2.10 package for R 3.0.1 with default settings. Genes expressed in less than 5 samples were removed, leaving 23460 genes. After transforming the normalized counts with the inverse hyperbolic sine, the genes with the 5% weakest coefficients of variation over the samples were removed, leaving 22287 genes for further analyses. Differential expression was calculated as mild drought minus control normalized expression values.

Cluster affinity search technique (CAST)

The CAST clustering, as described by (Ben-Dor et al., 1999), was performed on the normalized differential expression values of all 89 accessions. The clustering was executed by using the CAST implementation in Multiple Array Viewer (MeV) version 4.8 (Saeed et al., 2003). Pearson correlation was selected as a distance metric, the threshold was set at 0.7.

Stress predictors

Classification

A two-class classification was performed. Herein, the samples under stress were seen as the positive class, while the control samples were set as the negative class. Classification was performed by support vector machines (SVM). SVMs are large margin classifiers and rely on kernel functions to solve non-linear separability problems in a higher feature space. By transforming the data to a higher feature space, the classification problem can be solved using a linear model (Cortes and Vapnik, 1995; Hastie et al., 2005). In this study, a linear kernel was considered for classification. The software package scikit-learn v14 was used as modeling framework (Pedregosa et al., 2012). The SVM cost parameter C was optimized in the interval [2^-5, 2^15] with a step size of 2^2 (Chang and Lin, 2011). To ensure generalization and obtain an unbiased performance estimation, a stratified nested ten-fold cross-validation was performed (Kohavi, 1995; Varma and Simon, 2006). An inner cross-validation is used for model optimization, while an outer cross-validation is performed solely for performance estimation. Given the number of positive and negative samples in each outer test fold, a confusion matrix was generated to assess the performance of the classification model. Parameter optimization and performance estimation was done by the F-score, which is the harmonic mean between sensitivity and precision:

$$F = (2 \times \text{sensitivity} \times \text{precision})/(\text{sensitivity} + \text{precision})$$

$$\text{Sensitivity} = (\#\text{true positives})/(\#\text{true positives} + \#\text{false negatives})$$

$$\text{Precision} = (\#\text{true positives})/(\#\text{true positives} + \#\text{false positives})$$

Final performance estimation was calculated as the average score over the different outer cross-validation folds. Next to the F-score, the area under the receiver operating characteristics curve (AUC) was also calculated (Fawcett, 2006). The AUC can be interpreted as the probability of ranking a randomly chosen positive sample higher than randomly chosen negative sample. An

AUC value of 1 implies a perfect separation of positive and negative samples, while a value of 0.5 implies random classification.

Feature selection

To reduce the gene space of the two data sets and decrease the complexity of the classification models, we applied a filter feature selection technique (Saeys et al., 2007). Filter methods deal with the intrinsic properties of the data. ANOVA was chosen as a first step for reducing the gene space. A one-way ANOVA was calculated using stress treatment as a factor (R version 3.1.0). After multiple hypothesis testing corrections using the false discovery rate, significance was determined based on a threshold of 0.05.

Co-differential expression analysis

Co-differential expression analysis was done on the normalized differential expression values of the stress predictor genes. The Pearson correlation of the differential expression in 89 accessions was calculated between each pair of genes. Genes were considered to be co-differentially expressed when the Pearson correlation was at least 0.7. Visualisation of the output was done in Cytoscape version 3.1.0 (Shannon et al., 2003).

GO enrichment

All gene ontology (GO) enrichment analyses were conducted by using the online tool PLAZA v. 3.0 (Proost et al., 2015).

GWAS

The genome wide association mapping was performed on the normalized mean values of the different phenotypes in control and mild drought conditions. The mapping panel consisted of 91 accessions for which genotypic data could be obtained. The genotype data was based on whole genome sequencing data obtained within the 1001 genomes project (www.1001genomes.org) and covered ±4,000,000 SNPs. For accessions that were not sequenced, genotype information was imputed based on the 250k SNP chip data (Horton et al., 2012). Imputations of genotypic information are known to have negligible effect on the outcome of GWAS studies (Cao et al., 2011). Markers with a minor allele frequency below 5% were excluded from the analysis. The GWAS was performed by using a multi-trait mixed model (MTMM) as described by (Korte et al., 2012).

$$y \sim ß_0 + ß_1 E + ß_2 G + ß_3 GE + K + \varepsilon$$

Where y is a vector of phenotype values for n accessions, E is the treatment factor (control or mild drought stress), G the homozygous genotype of the respective marker, GE the genotype by treatment interaction. K denotes the kinship matrix that is included as a random factor to correct for population structure. The kinship matrix is calculated using the genotype information as an identical-by-state matrix, where relatedness is based on shared alleles, as discussed by (Kang et al., 2008). ß denotes the respective regression coefficient of each term.

To retrieve associations with the differential response against mild drought stress, the full model was tested against the model without GE ($ß_3 = 0$). A multiple testing correction was applied by dividing the 0.05 significance threshold by the number of markers with a minor allele frequency of at least 5%. For leaf 3 area at proliferation this resulted in a threshold of 7.53 –log10. Peak selection for the other phenotypes was done visually, resulting in following –log10 p-value cut-offs: 4.28 (rosette area), 5.46 (pavement cell area), 4.71 (pavement cell number), 4.42 (stomatal index), 5.26 (leaf 3 area at maturity). All analysis were conducted using a custom R-script (R version 3.1).

Pseudo-heritabilities were calculated according:

$$h = \frac{Var(G)}{Var(G) + Var(E)}$$

Var(G) and Var(E) are, respectively, the genetic variation and environmental variation as the variances estimated from the MTMM.

eGWAS

The eGWAS analysis was conducted with the same multi-trait mixed model as used for the GWAS analyses. As phenotype values, the gene expression data (log2 transcripts per million) was used. The trait-specific test was performed by testing the full model against the model without GE ($ß_3 = 0$). For the common test, the model without GE ($ß_3 = 0$) was tested against the model without G nor GE ($ß_2 = 0$ and $ß_3 = 0$). A fixed p-value cut-off for all genes was set at 10^{-7}. To create a high-confidence dataset, a selection was conducted against inflated models by calculating the genomic inflation factor (λ) (Devlin and Roeder, 1999) and performing the D'Agostino test for normality (D'Agostino, 1986). The genomic inflation factor was calculated as the ratio between the median of the p-values of the association of the markers with the trait (gene expression) and their expected median (0.5). Models with $\lambda = 1 \pm 0.1$ were considered to be non-inflated. The D'Agostino test on the gene expression values was performed as implemented in the R-package 'moments' (version 0.14). Genes with a p-value > 0.05 for the D'Agostino test in both treatments were considered to have normal distributed expression values and were subsequently selected as non-inflated. All analysis were conducted in R (version 3.1).

Supplementary tables

Accession	Region	Country
Ag-0	Western Europe	FRA
An-1	Western Europe	BEL
Bl-1	Southern Europe	ITA
Blh-1	Eastern Europe	CZE
Bor-1	Moravia	CZE
Bor-4	Moravia	CZE
Br-0	Moravia	CZE
Bur-0	Northern Europe	IRL
C24	Southern Europe	POR
Can-0	Southern Europe	ESP
CIBC-17	Northern Europe	UK
CIBC-5	Northern Europe	UK
Col-0	Midwest	USA
Cvi-0	Macaronesia	CPV
Del-10	Eastern Europe	RS
Edi-0	Northern Europe	UK
Ei-2	Western Europe	GER
Est-1	Northern Europe	EE
Fei-0	Southern Europe	POR
Ga-0	Western Europe	GER
Got-22	Western Europe	GER
Got-7	Western Europe	GER
Gu-0	Western Europe	GER
Gy-0	Western Europe	FRA
HR-10	Northern Europe	UK
HR-5	Northern Europe	UK
ICE112	Southern Europe	ITA
ICE138	Central Asia	RUS
ICE153	Central Asia	UZ
ICE163	Central-Europe	ITA
ICE61	Central Asia	RUS
ICE75	Central Asia	RUS
ICE97	Southern Europe	ITA
Jm-0	Eastern Europe	CZE
Kas-2	South Asia	IND
Kin-0	Midwest	USA
Knox-10	Indiana	USA
Knox-18	Indiana	USA
Kondara	Central Asia	TJK
Kz-1	Central Asia	KAZ
Kz-9	Central Asia	KAZ
Ler-1	Western Europe	GER
Lp2-2	Moravia	CZE
Lp2-6	Moravia	CZE
Lz-0	Western Europe	FRA
Mr-0	Southern Europe	ITA
Mrk-0	Western Europe	GER
Ms-0	Northern Europe	RUS
Mt-0	Cyrenaica (also Kyrenaika)	LIB
Mz-0	Western Europe	GER
N13	Northern Europe	RUS
Nd-1	Central Europe	SUI
NFA-10	Northern Europe	UK
NFA-8	Northern Europe	UK
Nok-3	Western Europe	NED
Omo2-1	S Sweden	SWE
OMo2-3	S Sweden	SWE
Oy-0	Northern Europe	NOR
Pna-10	Midwest	USA
Pna-17	Midwest	USA
Pro-0	Southern Europe	ESP
Pu2-23	Eastern Europe	CZE
Pu2-7	Eastern Europe	CZE
Ren-1	Western Europe	FRA
Ren-11	Western Europe	FRA
Rmx-A02	Midwest	USA
Rmx-A180	Midwest	USA
RRS-10	Midwest	USA
RRS-7	Midwest	USA
Sap-0	Eastern Europe	CZE
Se-0	Southern Europe	ESP
Sha	Central-Europe	TJK
Sorbo	Central Asia	TJK
Sp-0	Western Europe	GER
Spr1-6	C Sweden	SWE
Sq-1	Northern Europe	UK
Sq-8	Northern Europe	UK
St-0	C Sweden	SWE
Ta-0	Eastern Europe	CZE
Tamm-27	Northern Europe	FIN
Te-0	Northern Europe	FIN
Ts-1	Southern Europe	ESP
Ts-5	Southern Europe	ESP
Tsu-1	Eastern Asia	JPN
Ull2-3	S Sweden	SWE
Uod-1	Western Europe	AUT
Uod-7	Western Europe	AUT
Wa-1	Eastern Europe	POL
Walhasb4	Central-Europe	GER
Wei-0	Western Europe	SUI
Ws-0	Eastern Europe	RUS
Ws-2	Eastern Europe	RUS
Wt-5	Western Europe	GER
Yeg-1	Central Asia	AM
Yo-0	Northern California	USA
Zdr-1	Moravia	CZE
Zdr-6	Moravia	CZE

Supplementary Table 1: Geographic info on the 98 accessions used for phenotyping.

Peak	Gene	Gene Function
1	AT1G61260	unknown
1	AT1G61270	Transmembrane amino acid transporter family protein
1	AT1G61275	U12 small nucleolar RNA (U12)
1	AT1G61280	Phosphatidylinositol N-acetylglucosaminyltransferase,
1	AT1G61290	SYNTAXIN OF PLANTS 124 (SYP124)
1	AT1G61300	LRR and NB-ARC domains-containing disease resistance protein
1	AT1G61667	unknown
1	AT1G61670	Lung seven transmembrane receptor family protein
1	AT1G61680	TERPENE SYNTHASE 14 (TPS14)
1	AT1G61688	Defensin-like (DEFL) family protein
1	AT1G61690	phosphoinositide binding
1	AT1G61700	RNA polymerases N / 8 kDa subunit
1	AT1G62020	Coatomer, alpha subunit
1	AT1G62030	Cysteine/Histidine-rich C1 domain family protein
1	AT1G62035	microRNA171C (MIR171c)
1	AT1G62040	autophagy 8c (ATG8C)
1	AT1G62045	BEST Arabidopsis thaliana protein match is: ankyrin repeat family protein
1	AT1G62050	Ankyrin repeat family protein
2	AT3G60150	unknown
2	AT3G60160	ATP-binding cassette C9 (ABCC9), multidrug resistance-associated protein 9 (MRP9)
2	AT3G60176	other RNA
2	AT3G60180	P-loop containing nucleoside triphosphate hydrolases superfamily protein
2	AT3G60190	ARABIDOPSIS DYNAMIN-LIKE 4 (ADL4), ENHANCED DISEASE RESISTANCE 3 (EDR3)
2	AT3G60200	unknown
2	AT3G60210	GroES-like family protein
3	AT5G27600	LONG-CHAIN ACYL-COA SYNTHETASE 7 (LACS7)
3	AT5G27606	unknown
3	AT5G27610	ALWAYS EARLY 1 (ALY1)
4	AT5G28500	unknown
4	AT5G28510	BETA GLUCOSIDASE 24 (BGLU24)

Supplementary Table 2: Genes underlying the peaks (1-4) of SNPs associated with leaf 3 area at maturity. The selected peaks are visualized in **Figure 4**B.

Gene	Name	Number of Accessions	Differential Expression Upon Mild Drought
AT2G39800	*P5CS1*	84	Up-regulated
AT1G08650	*PPCK1*	81	Up-regulated
AT2G46680	*HB7*	80	Up-regulated
AT3G11410	*PP2CA/AHG3*	81	Up-regulated
AT5G64570	*XYL4*	80	Up-regulated
AT2G25625	unknown	80	Up-regulated
AT3G57520	*SIP2*	80	Down-regulated
AT1G69260	*AFP1*	83	Up-regulated
AT1G60960	*IRT3*	80	Up-regulated
AT2G03090	*EXPA15*	81	Up-regulated
AT5G53390	*FOP1*	84	Down-regulated

Supplementary Table 3: The eleven 'signature' genes with similar transcription responses to mild drought stress in at least 80 accessions. Number of accessions indicates for how many accessions the differential expression upon mild drought stress had the same direction.

Supplementary Table 4

Due to the large size of this table it is provided electronically via:
https://floppy.psb.ugent.be/public.php?service=files&t=11766ffffd6a153806ab76aeba1dc49c

Supplementary Table 4: CAST clusters. All genes of the CAST clusters containing the eleven 'signature' genes.

Supplementary Table 5

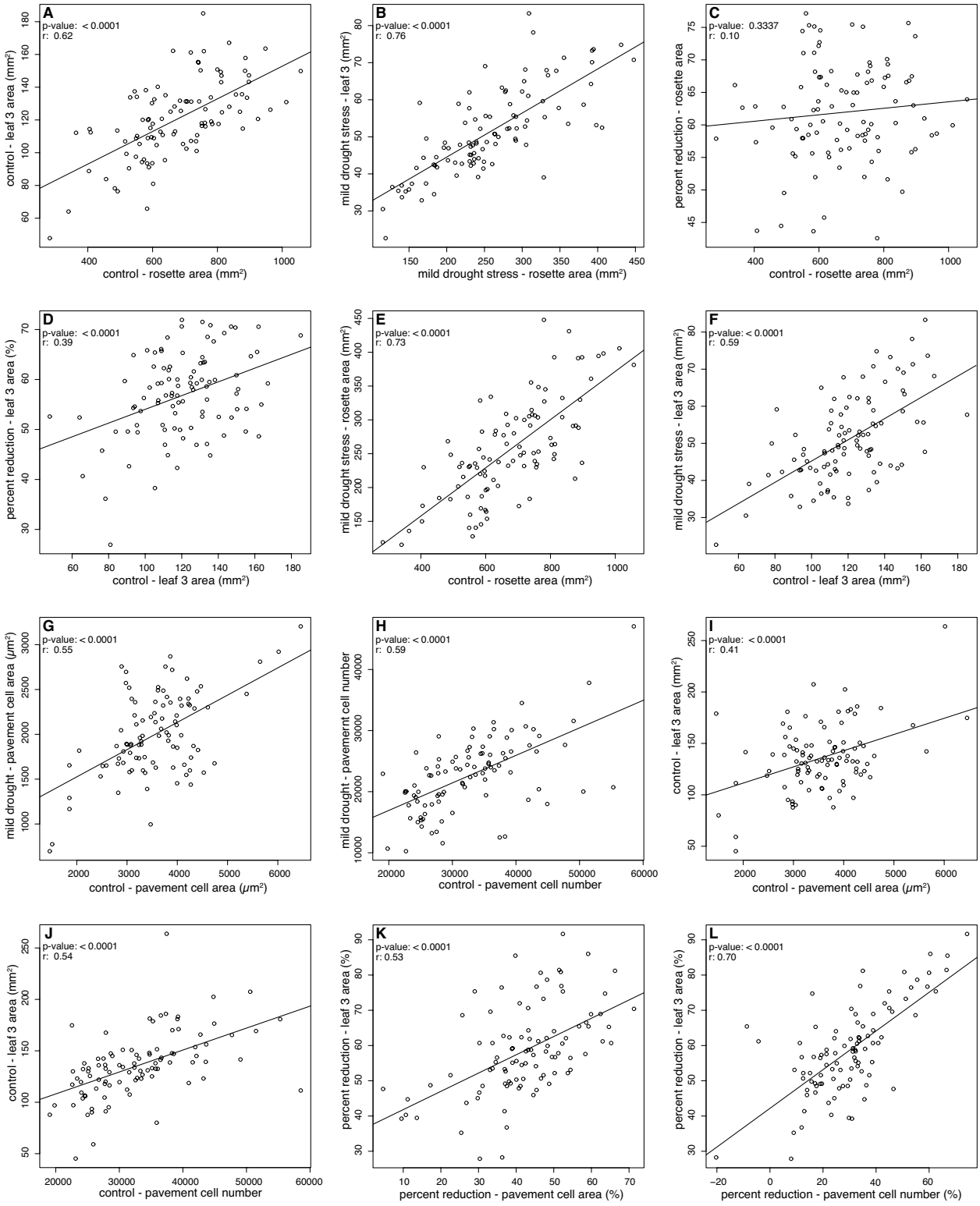
Due to the large size of this table it is provided electronically via:
https://floppy.psb.ugent.be/public.php?service=files&t=9f88b2881964f011b9caaefd78f33d58

Supplementary Table 5: Overlap between stress predictors and common drought genes

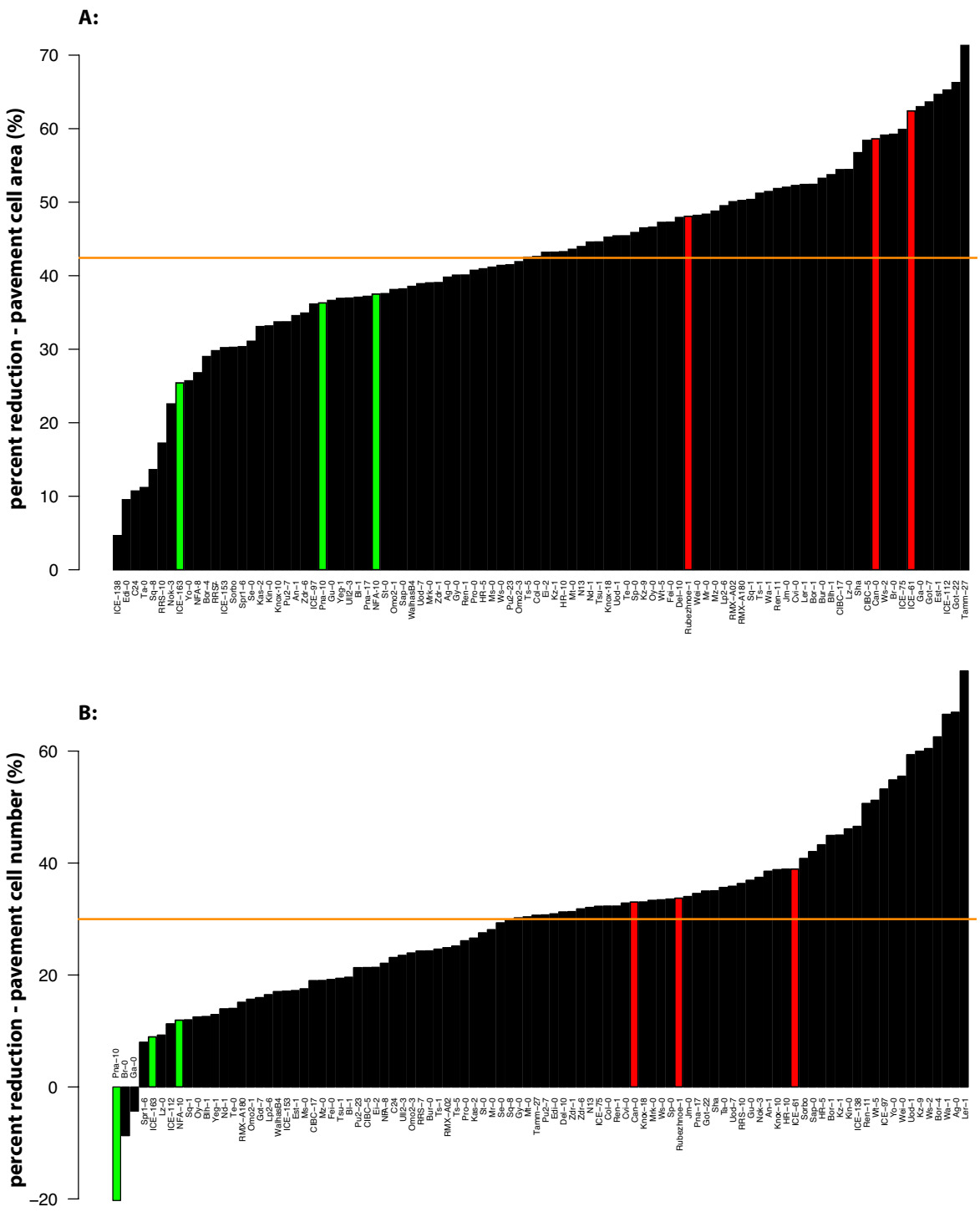
Gene	Gene Name/Function
AT2G41240	BASIC HELIX-LOOP-HELIX PROTEIN 100 (BHLH100)
AT3G56970	BASIC HELIX-LOOP-HELIX PROTEIN 38 (BHLH38)
AT3G56980	BASIC HELIX-LOOP-HELIX PROTEIN 39 (BHLH39)
AT5G61590	AP2 domain-containing transcription factor family
AT1G29395	COLD REGULATED 414 THYLAKOID MEMBRANE 1 (COR414-TM1)
AT1G47128	RESPONSIVE TO DEHYDRATION 21 (RD21)
AT1G69270	RECEPTOR-LIKE PROTEIN KINASE 1 (RPK1)
AT2G18050	HISTONE H1-3 (HIS1-3)
AT2G35300	late embryogenesis abundant group 1 domain-containing protein
AT3G02480	ABA-responsive protein-related
AT3G14440	NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 3 (NCED3)
AT3G62590	lipase class 3 family protein
AT3G63010	GA INSENSITIVE DWARF1B (GID1B)
AT4G02280	SUCROSE SYNTHASE 3 (SUS3)
AT5G06760	late embryogenesis abundant group 1 domain-containing protein
AT5G15970	COLD-RESPONSIVE 6 (COR6)
AT5G52300	LOW-TEMPERATURE-INDUCED 65 (LTI65)
AT5G66400	RESPONSIVE TO ABA 18 (RAB18)

Supplementary Table 6: Stress predictors that are not among the common drought genes (Clauw et al., 2015) and are within the GO category 'response to water deprivation'.

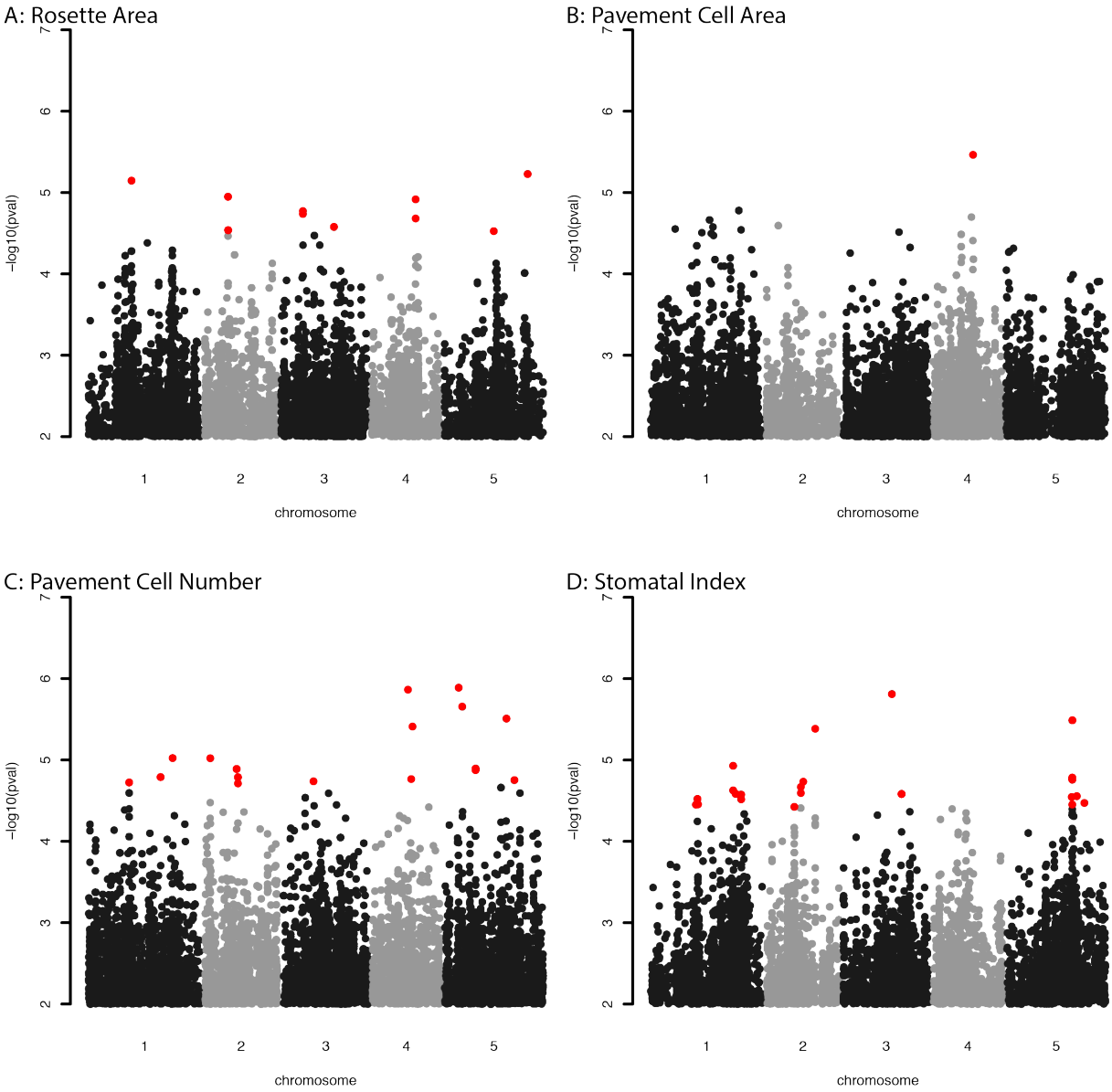
Supplementary figures



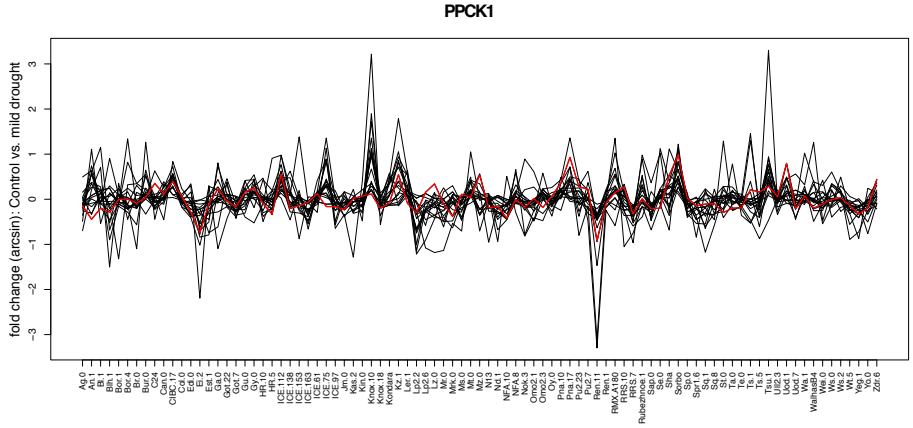
Supplementary Figure 1: Correlations of different phenotypic measurements in 98 accessions. A: leaf 3 area vs. rosette area in control conditions. B: leaf 3 area vs. rosette area in mild drought conditions at maturity. C: Percent reduction of rosette area vs. rosette area in control conditions. D: Percent reduction under mild drought stress of leaf 3 area vs. rosette area in control conditions. E: Rosette area under mild drought vs. control conditions. F: Leaf 3 area under mild drought vs. control conditions. G: Pavement cell area under mild drought vs. control conditions. H: Pavement cell number under mild drought vs. control conditions. I: leaf 3 area vs. pavement cell area in control conditions. J: leaf 3 area vs. pavement cell number in control conditions. K: Percent reduction of leaf 3 area vs. pavement cell area. L: Percent reduction of leaf 3 area vs. pavement cell area. Percent Reductions are the reductions under mild drought stress, relative to the control conditions. P-values give the significance of the correlations, r is the Pearson correlation coefficient.



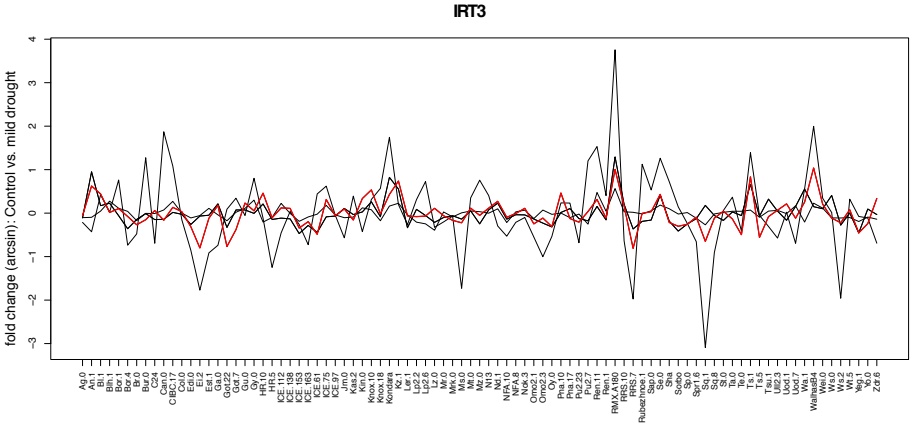
Supplementary Figure 2: Cellular response to mild drought stress of tolerant (green) and sensitive (red) accessions. A: Percent reductions of pavement cell area for 96 accessions. B: Percent reductions of pavement cell number for 96 accessions. Due to technical reasons there was no cellular data available for Kondara and Lp2-2. The average percentage reduction is indicated by the orange line.



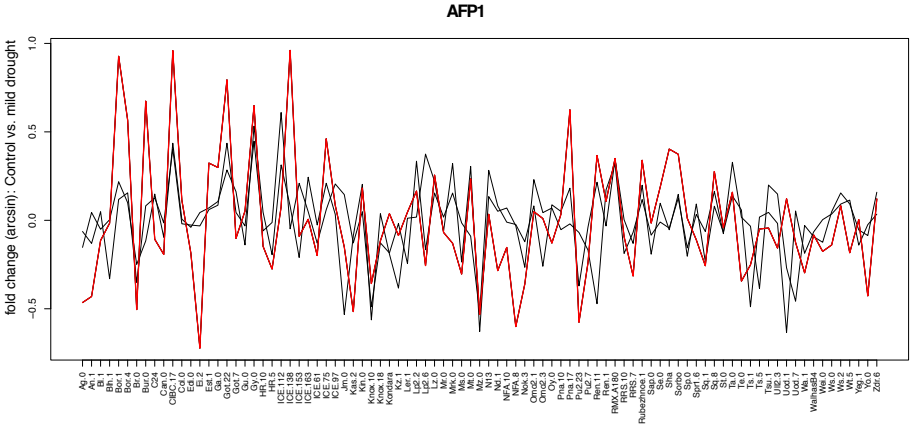
Supplementary Figure 3: Manhattan plots showing chosen SNPs (red) for gene selection. A: association of SNPs to differential response to mild drought of rosette area. B: association of SNPs to differential response to mild drought of pavement cell area. C: association of SNPs to differential response to mild drought of pavement cell number. D: association of SNPs to differential response to mild drought of stomatal index.



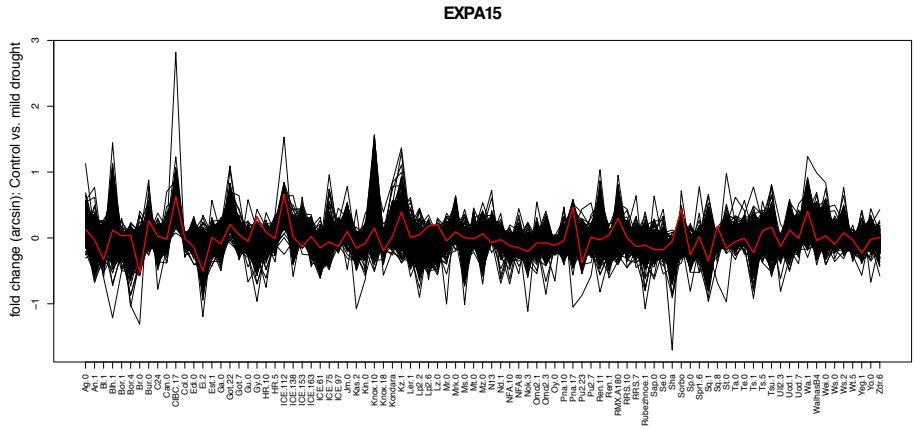
Supplementary Figure 4: The expression fold changes (control vs. mild drought; arcsin transformed) of the CAST cluster to which PPCK1 (red) belongs.



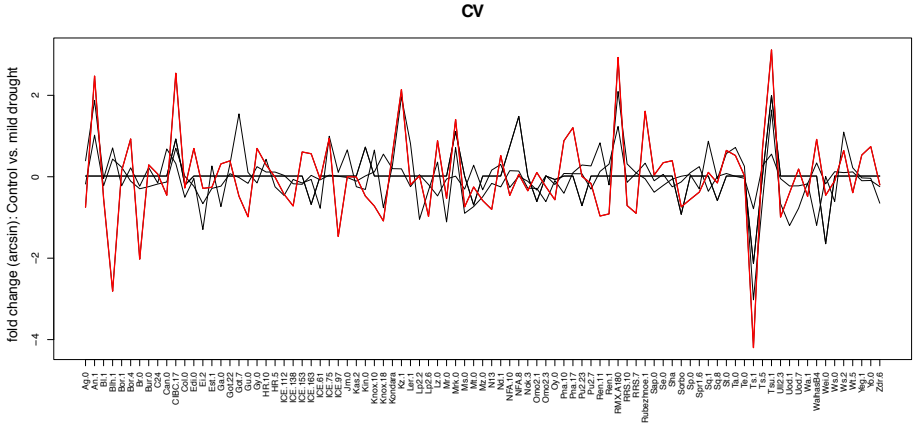
Supplementary Figure 5: The expression fold changes (control vs. mild drought; arcsin transformed) of the CAST cluster to which IRT3 (red) belongs.



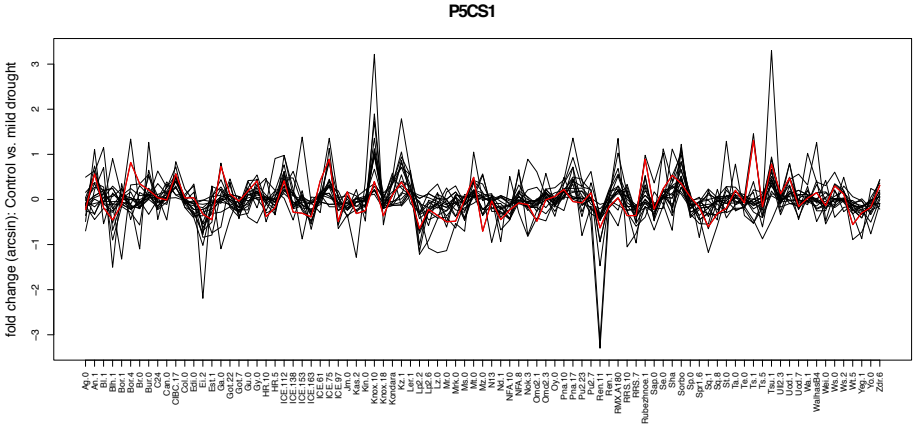
Supplementary Figure 6: The expression fold changes (control vs. mild drought; arcsin transformed) of the CAST cluster to which AFP1 (red) belongs.



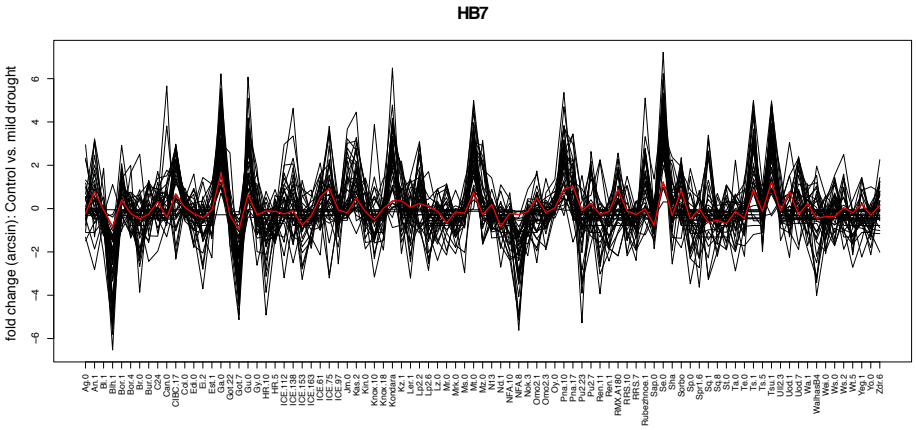
Supplementary Figure 7: The expression fold changes (control vs. mild drought; arcsin transformed) of the CAST cluster to which EXPA15 (red) belongs.



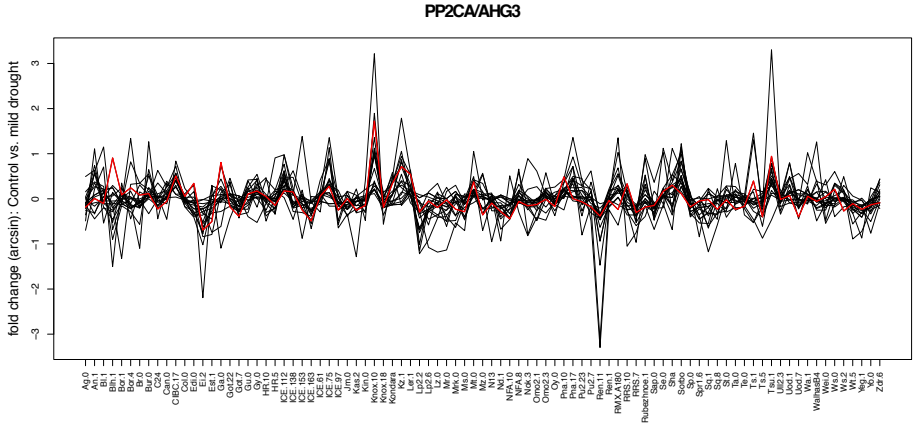
Supplementary Figure 8: The expression fold changes (control vs. mild drought; arcsin transformed) of the CAST cluster to which CV (red) belongs.



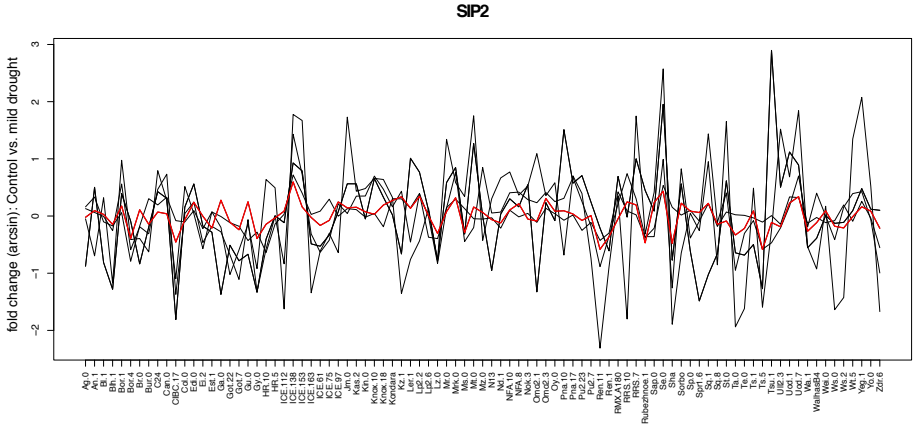
Supplementary Figure 9: The expression fold changes (control vs. mild drought; arcsin transformed) of the CAST cluster to which P5CS1 (red) belongs.



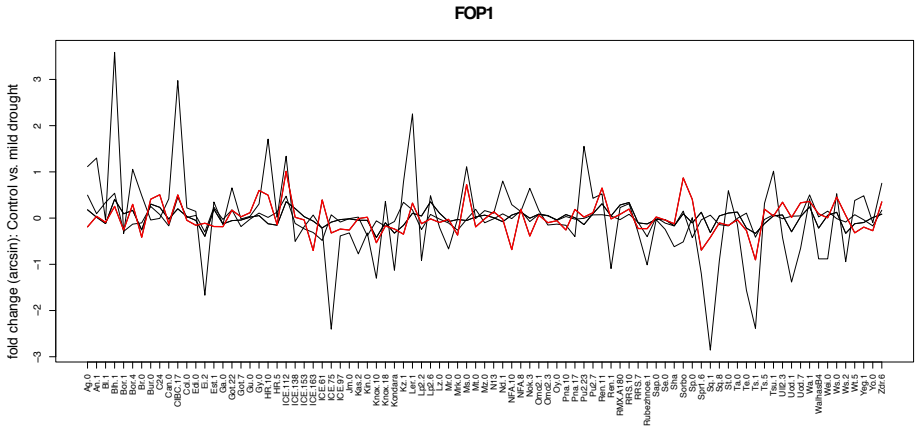
Supplementary Figure 10: The expression fold changes (control vs. mild drought; arcsin transformed) of the CAST cluster to which HB7 (red) belongs.



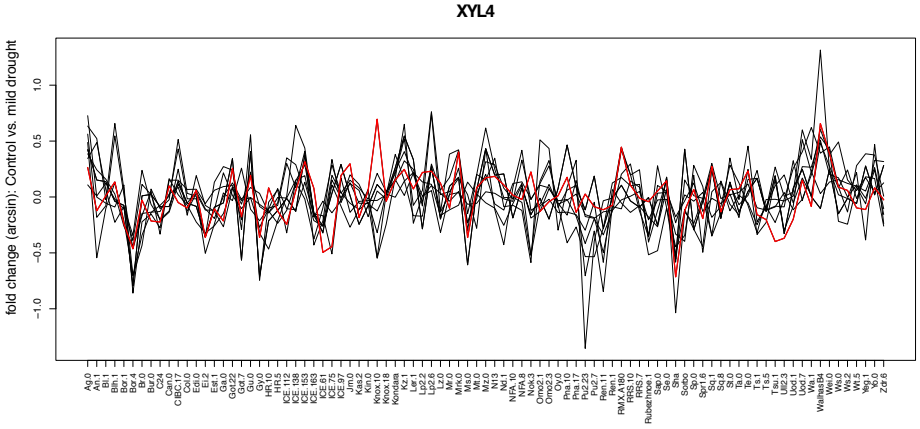
Supplementary Figure 11: The expression fold changes (control vs. mild drought; arcsin transformed) of the CAST cluster to which PP2CA/AHG3 (red) belongs.



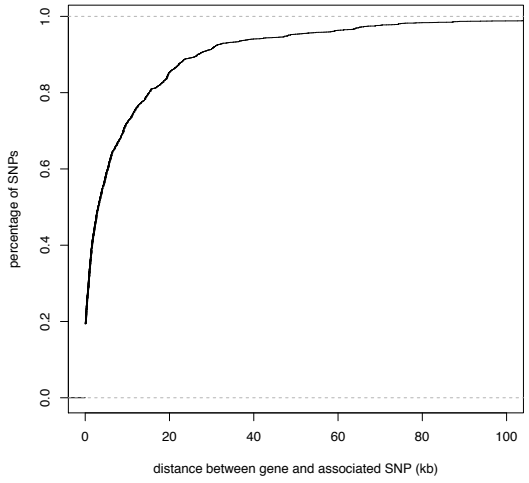
Supplementary Figure 12: The expression fold changes (control vs. mild drought; arcsin transformed) of the CAST cluster to which SIP2 (red) belongs.



Supplementary Figure 13: The expression fold changes (control vs. mild drought; arcsin transformed) of the CAST cluster to which FOP1 (red) belongs.



Supplementary Figure 14: The expression fold changes (control vs. mild drought; arcsin transformed) of the CAST cluster to which XYL4 (red) belongs.



Supplementary Figure 15: Cummulative percentage of SNPs over the distance to the transcription start site of the gene of which they associate with the expression, independent of treatment.

References

Achard P, Gusti A, Cheminant S, Alioua M, Dhondt S, Coppens F, Beemster GTS, Genschik P (2009) Gibberellin signaling controls cell proliferation rate in Arabidopsis. Curr Biol 19: 1188–1193

Aguirrezabal L, Bouchier-Combaud S, Radziejwoski A, Dauzat M, Cookson SJ, Granier C (2006) Plasticity to soil water deficit in Arabidopsis thaliana: dissection of leaf development into underlying growth dynamic and

cellular variables reveals invisible phenotypes. Plant Cell Environ 29: 2216–2227

Ahn CS, Han J-A, Pai H-S (2013) Characterization of in vivo functions of *Nicotiana benthamiana* RabE1. Planta 237: 161–172

Allocco DJ, Kohane IS, Butte AJ (2004) Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics 5: 18

Anastasiou E, Kenz S, Gerstung M, MacLean D, Timmer J, Fleck C, Lenhard M (2007) Control of Plant Organ Size by KLUH/CYP78A5-Dependent Intercellular Signaling. Dev Cell 13: 843–856

Anders S, Pyl PT, Huber W (2014) HTSeq–A Python framework to work with high-throughput sequencing data. bioRxiv. doi: 10.1101/002824

Andriankaja M, Dhondt S, De Bodt S, Vanhaeren H, Coppens F, De Milde L, Mühlenbock P, Skirycz A, Gonzalez N, Beemster GTS, et al (2012) Exit from proliferation during leaf development in *Arabidopsis thaliana*: a not-so-gradual process. Dev Cell 22: 64–78

Andriankaja ME, Danisman S, Mignolet-Spruyt LF, Claeys H, Kochanke I, Vermeersch M, De Milde L, De Bodt S, Storme V, Skirycz A, et al (2014) Transcriptional coordination between leaf cell differentiation and chloroplast development established by TCP20 and the subgroup Ib bHLH transcription factors. Plant Mol Biol 85: 233–245

Angelovici R, Lipka AE, Deason N, Gonzalez-Jorge S, Lin H, Cepela J, Buell R, Gore MA, Dellapenna D (2013) Genome-wide analysis of branched-chain amino acid levels in Arabidopsis seeds. Plant Cell 25: 4827–4843

Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, et al (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627–631

Ausin I, Greenberg MVC, Simanshu DK, Hale CJ, Vashisht AA, Simon SA, Lee T-F, Feng S, Espanola SD, Meyers BC, et al (2012) INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in Arabidopsis. Proc Natl Acad Sci USA 109: 8374–8381

Baerenfaller K, Massonnet C, Walsh S, Baginsky S, Bühlmann P, Hennig L, Hirsch-Hoffmann M, Howell KA, Kahlau S, Radziejwoski A, et al (2012) Systems-based analysis of Arabidopsis leaf growth reveals adaptation to water deficit. Mol Syst Biol 8: 606

Balazadeh S, Siddiqui H, Allu AD, Matallana-Ramirez LP, Caldana C, Mehrnia M, Zanor M-I, Köhler B, Mueller-Roeber B (2010) A gene regulatory network controlled by the NAC transcription factor ANAC092/AtNAC2/ORE1 during salt-promoted senescence. Plant J 62: 250–264

Baranauskė S, Mickutė M, Plotnikova A, Finke A, Venclovas Č, Klimašauskas S, Vilkaitis G (2015) Functional mapping of the plant small RNA methyltransferase: HEN1 physically interacts with HYL1 and DICER-LIKE 1 proteins. Nucleic Acids Res 43: 2802–2812

Bardou F, Ariel F, Simpson CG, Romero-Barrios N, Laporte P, Balzergue S, Brown JWS, Crespi M (2014) Long Noncoding RNA Modulates Alternative Splicing Regulators in Arabidopsis. DEVCEL 30: 166–176

Barriere A, Gordon KL, Ruvinsky I (2011) Distinct Functional Constraints Partition Sequence Conservation in a cis-Regulatory Element. PLoS Genet. doi: 10.1371/journal.pgen.1002095

Beemster G, Fiorani F, Inze D (2003) Cell cycle: the key to plant growth control? Trends Plant Sci 8: 154-158

Beemster GTS, Mironov V, Inzé D (2005) Tuning the cell-cycle engine for improved plant performance. Curr Opin Biotechnol 16: 142–146

Ben-Dor A, Shamir R, Yakhini Z (1999) Clustering gene expression patterns. J Comp Biol 6: 281–297

Bergmann DC, Sack FD (2007) Stomatal development. Annu Rev Plant Biol 58: 163–181

Bischoff V, Nita S, Neumetzler L, Schindelasch D, Urbain A, Eshed R, Persson S, Delmer D, Scheible WR (2010) TRICHOME BIREFRINGENCE and Its Homolog AT5G01360 Encode Plant-Specific DUF231 Proteins Required for Cellulose Biosynthesis in Arabidopsis. Plant Physiol 153: 590–602

Bouchabke O, Chang F, Simon M, Voisin R, Pelletier G, Durand-Tardif M (2008) Natural variation in Arabidopsis thaliana as a tool for highlighting differential drought responses. PLoS ONE 3: e1705

Brachi B, Faure N, Bergelson J, Cuguen J, Roux F (2013a) Genome-wide association mapping of flowering time in *Arabidopsis thaliana* in nature: genetics for underlying components and reaction norms across two successive years. Acta Botanica Gallica 160: 205–219

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F (2010) Linkage

and association mapping of *Arabidopsis thaliana* flowering time in nature. PLoS Genet 6: e1000940

Brachi B, Villoutreix R, Faure N, Hautekèete N, Piquot Y, Pauwels M, Roby D, Cuguen J, Bergelson J, Roux F (2013b) Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in Arabidopsis thaliana. Molecular Ecology 22: 4222–4240

Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci USA 102: 1572–1577

Bressan RA, Handa AK, Handa S, Hasegawa PM (1982) Growth and water relations of cultured tomato cells after adjustment to low external water potentials. Plant Physiol 70: 1303–1309

Bressan RA, Nelson DE, Iraki NM, LaRosa PC, Singh NK, Hasegawa PM, Carita NC (1990) Reduced cell expansion and changes in cell walls of plant cells adapted to NaCl.

Bulankova P, Riehs-Kearnan N, Nowack MK, Schnittger A, Riha K (2010) Meiotic progression in Arabidopsis is governed by complex regulatory interactions between SMG7, TDM1, and the meiosis I-specific cyclin TAM. Plant Cell 22: 3791–3803

Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, et al (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat Genet 43: 956–963

Capelson M, Liang Y, Schulte R, Mair W, Wagner U, Hetzer MW (2010) Chromatin-Bound Nuclear Pore Components Regulate Gene Expression in Higher Eukaryotes. Cell 140: 372–383

Carlsbecker A, Lee J-Y, Roberts CJ, Dettmer J, Lehesranta S, Zhou J, Lindgren O, Moreno-Risueno MA, Vatén A, Thitamadee S, et al (2010) Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. Nature 465: 316–321

Castaings L, Camargo A, Pocholle D, Gaudon V, Texier Y, Boutet-Mercey S, Taconnat L, Renou J-P, Daniel-Vedele F, Fernandez E, et al (2009) The nodule inception-like protein 7 modulates nitrate sensing and metabolism in Arabidopsis. Plant J 57: 426–435

Castillo EM, De Lumen BO, Reyes PS (1990) Raffinose synthase and galactinol synthase in developing seeds and leaves of legumes. Journal of Agricultural …. doi: 10.1021/jf00092a003

Chae K, Isaacs CG, Reeves PH, Maloney GS, Muday GK, Nagpal P, Reed JW (2012) Arabidopsis SMALL AUXIN UP RNA63 promotes hypocotyl and stamen filament elongation. Plant J 71: 684–697

Chang C-C, Lin C-J (2011) LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2: 27–27

Chardon F, Barthélémy J, Daniel-Vedele F, Masclaux-Daubresse C (2010) Natural variation of nitrate uptake and nitrogen use efficiency in Arabidopsis thaliana cultivated with limiting and ample nitrogen supply. J Exp Bot 61: 2293–2302

Chevalier C, Bourdon M, Pirrello J, Cheniclet C, Gévaudant F, Frangne N (2014) Endoreduplication and fruit growth in tomato: evidence in favour of the karyoplasmic ratio theory. J Exp Bot 65: 2731–2746

Cho HT, Cosgrove DJ (2000) Altered expression of expansin modulates leaf growth and pedicel abscission in Arabidopsis thaliana. Proc Natl Acad Sci USA 97: 9783–9788

Chou YH, Pogorelko G, Zabotina OA (2012) Xyloglucan Xylosyltransferases XXT1, XXT2, and XXT5 and the Glucan Synthase CSLC4 Form Golgi-Localized Multiprotein Complexes. Plant Physiol 159: 1355–1366

Claeys H, De Bodt S, Inzé D (2014) Gibberellins and DELLAs: central nodes in growth regulatory networks. Trends Plant Sci 19: 231–239

Claeys H, Inzé D (2013) The agony of choice: how plants balance growth and survival under water-limiting conditions. Plant Physiol 162: 1768–1779

Claeys H, Skirycz A, Maleux K, Inzé D (2012) DELLA signaling mediates stress-induced cell differentiation in Arabidopsis leaves through modulation of anaphase-promoting complex/cyclosome activity. Plant Physiol 159: 739–747

Clauw P, Coppens F, De Beuf K, Dhondt S, Van Daele T, Maleux K, Storme V, Clement L, Gonzalez N, Inzé D (2015) Leaf responses to mild drought stress in natural variants of Arabidopsis. Plant Physiol 167: 800–816

Cleland R (1971) Cell Wall Extension. Annu Rev Plant Physiol 22: 197–222

Coles JP, Phillips AL, Croker SJ, García-Lepe R, Lewis MJ, Hedden P (1999) Modification of gibberellin production and plant development in Arabidopsis by sense and antisense expression of gibberellin 20-oxidase genes. Plant J 17: 547–556

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20: 273–297

Cosgrove DJ (1987) Wall Relaxation and the Driving Forces for Cell Expansive Growth. Plant Physiol 84: 561–564

Cromer L, Heyman J, Touati S, Harashima H, Araou E, Girard C, Horlow C, Wassmann K, Schnittger A, De Veylder L, et al (2012) OSD1 Promotes Meiotic Progression via APC/C Inhibition and Forms a Regulatory Network with TDM and CYCA1;2/TAM. PLoS Genet 8: e1002865–14

D'Agostino RB (1986) Tests for the normal distribution. Goodness-of-fit techniques. doi: 10.1046/j.1365-2656.2002.00593.x/full

Davis J, Brandizzi F, Liepman AH, Keegstra K (2010) Arabidopsis mannan synthase CSLA9 and glucan synthase CSLC4 have opposite orientations in the Golgi membrane. Plant J 64: 1028–1037

Delker C, Pöschl Y, Raschke A, Ullrich K, Ettingshausen S, Hauptmann V, Grosse I, Quint M (2010) Natural variation of transcriptional auxin response networks in *Arabidopsis thaliana*. Plant Cell 22: 2184–2200

Deprost D, Yao L, Sormani R, Moreau M, Leterreux G, Nicolaï M, Bedu M, Robaglia C, Meyer C (2007) The Arabidopsis TOR kinase links plant growth, yield, stress resistance and mRNA translation. EMBO reports 8: 864–870

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997–1004

Di C, Yuan J, Wu Y, Li J, Lin H, Hu L, Zhang T, Qi Y, Gerstein MB, Guo Y, et al (2014) Characterization of stress-responsive lncRNAs in Arabidopsis thalianaby integrating expression, epigenetic and structural features. The Plant Journal 80: 848–861

Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Arcelus MG, Sekowska M, et al (2009) Common Regulatory Variation Impacts Gene Expression in a Cell Type-Dependent Manner. Science 325: 1246–1250

Dong Y, Wang C, Han X, Tang S, Liu S, Xia X, Yin W (2014) A novel bHLH transcription factor PebHLH35 from Populus euphratica confers drought tolerance through regulating stomatal development, photosynthesis and growth in Arabidopsis. Biochem Biophys Res Commun 450: 453–458

Drost DR, Benedict CI, Berg A, Novaes E, Novaes CRDB, Yu Q, Dervinis C, Maia JM, Yap J, Miles B, et al (2010) Diversification in the genetic architecture of gene expression and transcriptional networks in organ differentiation of Populus. Proc Natl Acad Sci USA 107: 8492–8497

Dubois M, Skirycz A, Claeys H, Maleux K, Dhondt S, De Bodt S, Vanden Bossche R, De Milde L, Yoshizumi T, Matsui M, et al (2013) ETHYLENE RESPONSE FACTOR6 acts as a central regulator of leaf growth under water-limiting conditions in Arabidopsis. Plant Physiol 162: 319–332

Efroni I, Blum E, Goldshmidt A, Eshed Y (2008) A Protracted and Dynamic Maturation Schedule Underlies Arabidopsis Leaf Development. Plant Cell 20: 2293–2306

Elbein AD, Pan YT, Pastuszak I, Carroll D (2003) New insights on trehalose: a multifunctional molecule. Glycobiology 13: 17R–27R

Ellison CE, Bachtrog D (2013) Dosage compensation via transposable element mediated rewiring of a regulatory network. Science 342: 846–850

Eloy NB, de Freitas Lima M, Van Damme D, Vanhaeren H, Gonzalez N, De Milde L, Hemerly AS, Beemster GTS, Inzé D, Ferreira PCG (2011) The APC/C subunit 10 plays an essential role in cell proliferation during leaf development. Plant J 68: 351–363

Eloy NB, Gonzalez N, Van Leene J (2012) SAMBA, a plant-specific anaphase-promoting complex/cyclosome regulator is involved in early development and A-type cyclin stabilization. doi: 10.1073/pnas.1211418109/-/DCSupplemental

Fawcett T (2006) An introduction to ROC analysis. Pattern Recognition Letters 27: 861–874

Ferjani A, Horiguchi G, Yano S, Tsukaya H (2007) Analysis of leaf development in fugu mutants of Arabidopsis reveals three compensation modes that modulate cell expansion in determinate organs. Plant Physiol 144: 988–999

Ferjani A, Ishikawa K, Asaoka M, Ishida M, Horiguchi G, Maeshima M, Tsukaya H (2014) Class III compensation, represented by KRP2overexpression, depends on V-ATPase activity in proliferative cells. Plant Signal Behav 8: e27204

Figueroa P, Léon G, Elorza A, Holuigue L, Araya A, Jordana X (2002) The four subunits of mitochondrial respiratory complex II are encoded by multiple nuclear genes and targeted to mitochondria in Arabidopsis thaliana. Plant Mol Biol 50: 725–734

Garcia ME, Lynch T, Peeters J, Snowden C, Finkelstein R (2008) A small plant-specific protein family of ABI five binding proteins (AFPs) regulates stress response in germinating Arabidopsis seeds and seedlings. Plant Mol Biol 67: 643–658

Gee MA, Hagen G, Guilfoyle TJ (1991) Tissue-specific and organ-specific expression of soybean auxin-responsive transcripts GH3 and SAURs. Plant Cell 3: 419–430

Geisler M, Nadeau J, Sack FD (2000) Oriented asymmetric divisions that generate the stomatal spacing pattern in arabidopsis are disrupted by the too many mouths mutation. Plant Cell 12: 2075–2086

Gifford ML, Banta JA, Katari MS, Hulsmans J, Chen L, Ristova D, Tranchina D, Purugganan MD, Coruzzi GM, Birnbaum KD (2013) Plasticity regulators modulate specific root traits in discrete nitrogen environments. PLoS Genet. doi: 10.1371/journal.pgen.1003760

Gonzalez N, De Bodt S, Sulpice R, Jikumaru Y, Chae E, Dhondt S, Van Daele T, De Milde L, Weigel D, Kamiya Y, et al (2010) Increased leaf size: different means to an end. Plant Physiol 153: 1261–1279

Gonzalez N, Vanhaeren H, Inzé D (2012) Leaf size control: complex coordination of cell division and expansion. Trends Plant Sci 17: 332–340

Green PB, Erickson RO, Buggy J (1971) Metabolic and physical control of cell elongation rate *in vivo* studies in Nitella. Plant Physiol 47: 423–430

Grundberg E, Adoue V, Kwan T, Ge B, Duan QL, Lam KCL, Koka V, Kindmark A, Weiss ST, Tantisira K, et al (2011) Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. PLoS Genet. doi: 10.1371/journal.pgen.1001279

Hadjiargyrou M, Delihas N (2013) The intertwining of transposable elements and non-coding RNAs. IJMS 14: 13307–13328

Hastie T, Tibshirani R, Friedman J (2005) The elements of statistical learning: data mining, inference and prediction. Springer series in statistics. 2009: Springer. 745.

Heyman J, Cools T, Vandenbussche F, Heyndrickx KS (2013) ERF115 controls root quiescent center cell division and stem cell replenishment. Science 342: 860-863

Hincha DK, Zuther E, Heyer AG (2003) The preservation of liposomes by raffinose family oligosaccharides during drying is mediated by effects on fusion and lipid phase transitions. Biochimica et Biophysica Acta (BBA) 162: 172-177

Hisamatsu T (2005) The involvement of gibberellin 20-oxidase genes in phytochrome-regulated petiole elongation of *Arabidopsis*. Plant Physiol 138: 1106–1116

Hollister JD, Smith LM, Guo Y-L, Ott F, Weigel D, Gaut BS (2011) Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and Arabidopsis lyrata. Proc Natl Acad Sci USA 108: 2322–2327

Hong S-Y, Kim O-K, Kim S-G, Yang M-S, Park C-M (2011) Nuclear import and DNA binding of the ZHD5 transcription factor is modulated by a competitive peptide inhibitor in Arabidopsis. J Biol Chem 286: 1659-1668

Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Muliyati NW, Platt A, Sperone FG, Vilhjálmsson BJ, et al (2012) Genome-wide patterns of genetic variation in worldwide. Nat Genet 44: 212–216

Hu Y, Poh HM, Chua N-H (2006) The Arabidopsis ARGOS-LIKE gene regulates cell expansion during organ growth. Plant J 47: 1–9

Hu Y, Xie Q, Chua N-H (2003) The Arabidopsis auxin-inducible gene ARGOS controls lateral organ size. Plant Cell 15: 1951–1961

Huang S, Raman AS, Ream JE, Fujiwara H, Cerny RE, Brown SM (1998) Overexpression of 20-oxidase confers a gibberellin-overproduction phenotype in Arabidopsis. Plant Physiol 118: 773–781

Huard-Chauveau C, Perchepied L, Debieu M, Rivas S, Kroj T, Kars I, Bergelson J, Roux F, Roby D (2013) An Atypical Kinase under Balancing Selection Confers Broad-Spectrum Disease Resistance in Arabidopsis. PLoS Genet. doi: 10.1371/journal.pgen.1003766

Hughes KA, Ayroles JF, Reedy MM, Drnevich JM, Rowe KC, Ruedi EA, Caceres CE, Paige KN (2006) Segregating variation in the transcriptome: Cis regulation and additivity of effects. Genetics 173: 1347–1364

Hur Y-S, Um J-H, Kim S, Kim K, Park H-J, Lim J-S, Kim W-Y, Jun SE, Yoon EK, Lim J, et al (2015) Arabidopsis thaliana homeobox 12 (ATHB12), a homeodomain-leucine zipper protein, regulates leaf growth by promoting cell expansion and endoreduplication. New Phytol 205: 316–328

Irshad M, Canut H, Borderies G, Pont-Lezica R, Jamet E (2008) A new picture of cell wall protein dynamics in

elongating cells of Arabidopsis thaliana: Confirmed actors and newcomers. BMC Plant Biology 8: 94–16

Ji L, Chen X (2012) Regulation of small RNA stability: methylation and beyond. Cell Res 22: 624–636

Jung J-H, Park C-M (2006) MIR166/165 genes exhibit dynamic expression patterns in regulating shoot apical meristem and floral development in Arabidopsis. Planta 225: 1327–1338

Kalve S, De Vos D, Beemster GTS (2014) Leaf development: a cellular perspective. Front Plant Sci 5: 362

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178: 1709–1723

Kapitonov VV, Jurka J (2007) Helitrons on a roll: eukaryotic rolling-circle transposons. Trends in Genetics 23: 521–529

Katori T, Ikeda A, Iuchi S, Kobayashi M, Shinozaki K, Maehashi K, Sakata Y, Tanaka S, Taji T (2010) Dissecting the genetic control of natural variation in salt tolerance of *Arabidopsis thaliana* accessions. J Exp Bot 61: 1125–1138

Kazama D, Kurusu T, Mitsuda N (2014) Involvement of elevated proline accumulation in enhanced osmotic stress tolerance in Arabidopsis conferred by chimeric repressor gene silencing technology. Plant signaling & Behavior doi: 10.4161/psb.28211

Kazama T, Ichihashi Y, Murata S, Tsukaya H (2010) The mechanism of cell cycle arrest front progression explained by a KLUH/CYP78A5-dependent mobile growth factor in developing leaves of Arabidopsis thaliana. Plant Cell Physiol 51: 1046–1054

Kenney AM, McKay JK, Richards JH (2014) Direct and indirect selection on flowering time, water-use efficiency (WUE, δ13C), and WUE plasticity to drought in Arabidopsis thaliana. Ecology and Evolution doi: 10.1002/ece3.1270

Keurentjes JJB, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJM, Vreugdenhil D, Koornneef M, Jansen RC (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proc Natl Acad Sci USA 104: 1708–1713

Kianian PMA, Kianian SF (2014) Mitochondrial dynamics and the cell cycle. Front Plant Sci. doi: 10.3389/fpls.2014.00222

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and linkage disequilibrium in Arabidopsis thaliana. Nat Genet 39: 1151–1155

Kim T-H, Böhmer M, Hu H, Nishimura N, Schroeder JI (2010) Guard cell signal transduction network: advances in understanding abscisic acid, CO2, and Ca2+ signaling. Annu Rev Plant Biol 61: 561–591

Kimbrough JM, Salinas-Mondragon R, Boss WF, Brown CS, Sederoff HW (2004) The fast and transient transcriptional network of gravity and mechanical stimulation in the Arabidopsis root apex. Plant Physiol 136: 2790–2805

Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. IJCAI

Kong Y, Zhu Y, Gao C, She W, Lin W, Chen Y, Han N, Bian H, Zhu M, Wang J (2013) Tissue-specific expression of SMALL AUXIN UP RNA41 differentially regulates cell expansion and root meristem patterning in Arabidopsis. Plant Cell Physiol 54: 609–621

Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M (2012) A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat Genet 44: 1066–1071

Kosma DK, Bourdenx B, Bernard A, Parsons EP, Lü S, Joubès J, Jenks MA (2009) The impact of water deficiency on leaf cuticle lipids of Arabidopsis. Plant Physiol 151: 1918–1929

Krizek BA (1999) Ectopic expression of AINTEGUMENTA in Arabidopsis plants results in increased growth of floral organs. Dev Genet 25: 224–236

Lake JA, Quick WP, Beerling DJ, Woodward FI (2001) Plant development: signals from mature to new leaves. Nature 411: 154–154

Lee BH, Ko J-H, Lee S, Lee Y, Pak J-H, Kim JH (2009) The Arabidopsis GRF-INTERACTING FACTOR gene family performs an overlapping function in determining organ size as well as multiple developmental properties. Plant Physiol 151: 655–668

Li C, Potuschak T, Colón-Carmona A, Gutiérrez RA, Doerner P (2005) Arabidopsis TCP20 links regulation of growth and cell division control pathways. Proc Natl Acad Sci USA 102: 12978–12983

Li F, Wu X, Lam P, Bird D, Zheng H, Samuels L, Jetter R, Kunst L (2008) Identification of the wax ester

synthase/acyl-coenzyme A: diacylglycerol acyltransferase WSD1 required for stem wax ester biosynthesis in Arabidopsis. Plant Physiol 148: 97–107

Li Y, Huang Y, Bergelson J, Nordborg M, Borevitz JO (2010) Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. PNAS 107: 21199–21204

Liu L, Adrian J, Pankin A, Hu J, Dong X, Korff von M, Turck F (2014a) Induced and natural variation of promoter length modulates the photoperiodic response of FLOWERING LOCUS T. Nat Commun 5: 4558

Liu W, Tai H, Li S, Gao W, Zhao M, Xie C, Li W-X (2014b) bHLH122 is important for drought and osmotic stress resistance in Arabidopsis and in the repression of ABA catabolism. New Phytol 201: 1192–1204

Lockhart JA (1965) An analysis of irreversible plant cell elongation. J Theor Biol 8: 264–275

Lorbiecke R, Sauter M (2002) Comparative analysis of PSK peptide growth factor precursor homologs. Plant Science 163: 321–332

Lu H, Zou Y, Feng N (2010) Overexpression of AHL20 Negatively Regulates Defenses in Arabidopsis. Journal of Integrative Plant Biology 52: 801–808

Lunn JE, Feil R, Hendriks J, Gibon Y, Morcuende R, Osuna D, Scheible WR, Carillo P, Hajirezaei MR, Stitt M (2006) Sugar-induced increases in trehalose 6-phosphate are correlated with redox activation of ADPglucose pyrophosphorylase and higher rates of starch synthesis in Arabidopsis thaliana. Biochem J 397: 139–148

Maggio A, Zhu J-K, Hasegawa PM, Bressan RA (2006) Osmogenetics: Aristotle to Arabidopsis. Plant Cell 18: 1542–1557

Marais Des DL, McKay JK, Richards JH, Sen S, Wayne T, Juenger TE (2012) Physiological Genomics of Response to Soil Drying in Diverse Arabidopsis Accessions. Plant Cell 24: 893–914

Markakis MN, Boron AK, Van Loock B, Saini K, Cirera S, Verbelen J-P, Vissenberg K (2013) Characterization of a Small Auxin-Up RNA (SAUR)-Like Gene Involved in Arabidopsis thaliana Development. PLoS ONE 8: e82596–13

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetJournal 17: pp. 10–12–12

Martín-Trillo M, Cubas P (2010) TCP genes: a family snapshot ten years later. Trends Plant Sci 15: 31–39

Matsubayashi Y, Ogawa M, Kihara H, Niwa M, Sakagami Y (2006) Disruption and overexpression of Arabidopsis phytosulfokine receptor gene affects cellular longevity and potential for growth. Plant Physiol 142: 45–53

Matsuda K, Riazi A (1981) Stress-induced osmotic adjustment in growing regions of barley leaves. Plant Physiol 68: 571–576

Mayfield JD, Paul A-L, Ferl RJ (2012) The 14-3-3 proteins of Arabidopsis regulate root growth and chloroplast development as components of the photosensory system. J Exp Bot 63: 3061–3070

McClure BA, Guilfoyle T (1989) Rapid Redistribution of Auxin-Regulated Rnas During Gravitropism. Science 243: 91–93

Meijón M, Satbhai SB, Tsuchimatsu T, Busch W (2014) Genome-wide association study using cellular traits identifies a new regulator of root development in Arabidopsis. Nat Genet 46: 77–81

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL (2003) Rapid evolution of male-biased gene expression in Drosophila. Proc Natl Acad Sci USA 100: 9894–9899

Meimoun P, Gousset-Dupont A, Lebouteiller B, Ambard-Bretteville F, Besin E, Lelarge C, Mauve C, Hodges M, Vidal J (2009) The impact of PEPC phosphorylation on growth and development of Arabidopsis thaliana: molecular and physiological characterization of PEPC kinase mutants. FEBS Letters 583: 1649–1652

Meyer RF, Boyer JS (1981) Osmoregulation, solute distribution, and growth in soybean seedlings having low water potentials. Planta 151: 482–489

Minic Z, Rihouey C, Do CT, Lerouge P, Jouanin L (2004) Purification and characterization of enzymes exhibiting beta-D-xylosidase activities in stem tissues of Arabidopsis. Plant Physiol 135: 867–878

Mizukami Y, Fischer RL (2000) Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. Proc Natl Acad Sci USA 97: 942–947

Moore JP, Vicré-Gibouin M, Farrant JM, Driouich A (2008) Adaptations of higher plant cell walls to water loss: drought vs desiccation. Physiol Plant 134: 237–245

Motose H, Iwamoto K, Endo S, Demura T, Sakagami Y, Matsubayashi Y, Moore KL, Fukuda H (2009) Involvement of phytosulfokine in the attenuation of stress response during the transdifferentiation of zinnia mesophyll cells into tracheary elements. Plant Physiol 150: 437–447

Muller J, Aeschbacher RA, Wingler A, Boller T, Wiemken A (2001) Trehalose and trehalase in Arabidopsis. Plant Physiol 125: 1086–1093

Olsson ASB, Engström P, Söderman E (2004) The homeobox genes ATHB12 and ATHB7 encode potential regulators of growth in response to water deficit in Arabidopsis. Plant Mol Biol 55: 663–677

Ori N, Cohen AR, Etzioni A, Brand A, Yanai O, Shleizer S, Menda N, Amsellem Z, Efroni I, Pekker I, et al (2007) Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato. Nat Genet 39: 787–791

Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC, Weigel D (2003) Control of leaf morphogenesis by microRNAs. Nature 425: 257–263

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al (2012) Scikit-learn: Machine Learning in Python. arXiv cs.LG:

Pereyra-Irujo GA, Velazquez L, Lechner L, Aguirrezabal L (2008) Genetic variability for leaf growth rate and duration under water deficit in sunflower: analysis of responses at cell, organ, and plant level. J Exp Bot 59: 2221–2232

Peters S, Egert A, Stieger B, Keller F (2010) Functional Identification of Arabidopsis ATSIP2 (At3g57520) as an Alkaline -Galactosidase with a Substrate Specificity for Raffinose and an Apparent Sink-Specific Expression Pattern. Plant Cell Physiol 51: 1815–1819

Peterson KM, Rychel AL, Torii KU (2010) Out of the mouths of plants: the molecular basis of the evolution and diversity of stomatal development. Plant Cell 22: 296–306

Pillitteri LJ, Dong J (2013) Stomatal development in Arabidopsis. The Arabidopsis Book 11: e0162

Proost S, Van Bel M, Vaneechoutte D, Van de Peer Y, Inzé D, Mueller-Roeber B, Vandepoele K (2015) PLAZA 3.0: an access point for plant comparative genomics. Nucleic Acids Res 43: D974–81

Quarrie SA, Jones HG (1977) Effects of Abscisic-Acid and Water Stress on Development and Morphology of Wheat. J Exp Bot 28: 192–203

Raser JM, O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. Science 304: 1811–1814

Rebollo R, Romanish MT, Mager DL (2012) Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes. Annu Rev Genet 46: 21–42

Ré DA, Capella M, Bonaventure G, Chan RL (2014) Arabidopsis AtHB7 and AtHB12 evolved divergently to fine tune processes associated with growth and responses to water stress. 14: 1–14

Risseeuw E, Venglat P, Xiang D, Komendant K, Daskalchuk T, Babic V, Crosby W, Datla R (2013) An Activated Form of UFO Alters Leaf Development and Produces Ectopic Floral and Inflorescence Meristems. PLoS ONE 8: e83807–12

Rodrigues A, Adamo M, Crozet P, Margalha L, Confraria A, Martinho C, Elias A, Rabissi A, Lumbreras V, Gonzalez-Guzman M, et al (2013) ABI1 and PP2CA phosphatases are negative regulators of Snf1-related protein kinase1 signaling in Arabidopsis. Plant Cell 25: 3871–3884

Rodriguez RE, Mecchia MA, Debernardi JM, Schommer C, Weigel D, Palatnik JF (2010) Control of cell proliferation in Arabidopsis thaliana by microRNA miR396. Development 137: 103–112

Rosas U, Cibrian-Jaramillo A, Ristova D, Banta JA, Gifford ML, Fan AH, Zhou RW, Kim GJ, Krouk G, Birnbaum KD, et al (2013) Integration of responses within and across Arabidopsis natural accessions uncovers loci controlling root systems architecture. Proc Natl Acad Sci USA 110: 15133–15138

Ruiqin Z, Chanhui L, Zheng-Hua Y (2010) Global Analysis of Direct Targets of Secondary Wall NAC Master Switches in Arabidopsis. Molecular Plant 3: 1087–1103

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, et al (2003) TM4: a free, open-source system for microarray data management and analysis. BioTechniques 34: 374–378

Saeys Y, Inza I, Larranaga P (2007) A review of feature selection techniques in bioinformatics. Bioinformatics 23: 2507–2517

Sakurai N, Akiyama M, Kuraishi S (1986) Irreversible Effects of Water-Stress on Growth and Stomatal Development in Cotyledons of Etiolated Squash Seedlings. Plant Cell Physiol 27: 1177–1185

Sampedro J, Cosgrove DJ (2005) The expansin superfamily. Genome Biol 6: 242

Saravitz DM, Pharr DM, Carter TE (1987) Galactinol Synthase Activity and Soluble Sugars in Developing Seeds of 4

Soybean Genotypes. Plant Physiol 83: 185–189

Sauter M (2015) Phytosulfokine peptide signalling. J Exp Bot 1–9

Schmuths H, Meister A, Horres R, Bachmann K (2004) Genome size variation among accessions of *Arabidopsis thaliana*. Ann Bot **93**: 317–321

Schröder F, Lisso J, Müssig C (2011) EXORDIUM-LIKE1 promotes growth during low carbon availability in Arabidopsis. Plant Physiol 156: 1620–1630

Schröder F, Lisso J, Müssig C (2012) Expression pattern and putative function of EXL1 and homologous genes in Arabidopsis. Plant Signal Behav 7: 22–27

Schulze S, Schäfer BN, Parizotto EA, Voinnet O, Theres K (2010) LOST MERISTEMS genes regulate cell differentiation of central zone descendants in Arabidopsis shoot meristems. The Plant Journal 64: 668–678

Schweizer P, Stein N (2011) Large-scale data integration reveals colocalization of gene functional groups with meta-QTL for multiple disease resistance in barley. Molecular Plant-Microbe Interactions 24: 1492–1501

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet 44: 825–830

Shabala S (2013) Learning from halophytes: physiological basis and strategies to improve abiotic stress tolerance in crops. Ann Bot 112: 1209–1221

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research 13: 2498–2504

Skirycz A, Inzé D (2010) More from less: plant growth under limited water. Curr Opin Biotechnol 21: 197–203

Skirycz A, Vandenbroucke K, Clauw P, Maleux K, De Meyer B, Dhondt S, Pucci A, Gonzalez N, Hoeberichts F, Tognetti VB, et al (2011) Survival and growth of Arabidopsis plants given limited water are not equal. Nat Biotechnol 29: 212–214

Slovak R, Goeschl C, Su X, Shimotani K, Shiina T, Busch W (2014) A Scalable Open-Source Pipeline for Large-Scale Root Phenotyping of Arabidopsis. Plant Cell 26: 2390–2403

Song JB, Gao S, Sun D, Li H, Shu XX, Yang ZM (2013) miR394 and LCR are involved in Arabidopsis salt and drought stress responses in an abscisic acid-dependent manner. BMC Plant Biology 13: 210

Söderman E, Mattsson J, Engström P (1996) The Arabidopsis homeobox gene ATHB-7 is induced by water deficit and by abscisic acid. Plant J 10: 375–381

Spartz AK, Lee SH, Wenger JP, Gonzalez N, Itoh H, Inzé D, Peer WA, Murphy AS, Overvoorde PJ, Gray WM (2012) The SAUR19 subfamily of SMALL AUXIN UP RNA genes promote cell expansion. The Plant Journal 70: 978–990

Staiger D, Brown JWS (2013) Alternative splicing at the intersection of biological timing, development, and stress responses. Plant Cell 25: 3640–3656

Stein RJ, Waters BM (2012) Use of natural variation reveals core genes in the transcriptome of iron-deficient *Arabidopsis thaliana* roots. J Exp Bot 63: 1039–1055

Stokes TL, Kunkel BN, Richards EJ (2002) Epigenetic variation in Arabidopsis disease resistance. Genes Dev 16: 171–182

Stransfeld L, Eriksson S, Adamski NM, Breuninger H, Lenhard M (2010) KLUH/CYP78A5 promotes organ growth without affecting the size of the early primordium. Plant Signal Behav 5: 982–984

Strizhov N, Abrahám E, Ökrész L, Blickling S, Zilberstein A, Schell J, Koncz C, Szabados L (1997) Differential expression of two P5CS genes controlling proline accumulation during salt-stress requires ABA and is regulated by ABA1, ABI1 and AXR2 in Arabidopsis. Plant J 12: 557–569

Ströher E, Wang X-J, Roloff N, Klein P, Husemann A, Dietz K-J (2009) Redox-dependent regulation of the stress-induced zinc-finger protein SAP12 in Arabidopsis thaliana. Molecular Plant 2: 357–367

Szabados L, Savouré A (2010) Proline: a multifunctional amino acid. Trends Plant Sci 15: 89–97

Taguchi N, Ishihara N, Jofuku A, Oka T, Mihara K (2007) Mitotic phosphorylation of dynamin-related GTPase Drp1 participates in mitochondrial fission. J Biol Chem 282: 11521–11529

Taiz L (1984) Plant Cell Expansion: Regulation of Cell Wall Mechanical Properties. Annu Rev Plant Physiol 35: 585–657

Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K (2002) Important

roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in Arabidopsis thaliana. Plant J 29: 417–426

Takeda S, Iwasaki A, Matsumoto N, Uemura T, Tatematsu K, Okada K (2013) Physical Interaction of Floral Organs Controls Petal Morphogenesis in Arabidopsis. Plant Physiol 161: 1242–1250

Tamura K, Fukao Y, Iwamoto M, Haraguchi T, Hara-Nishimura I (2010) Identification and characterization of nuclear pore complex components in Arabidopsis thaliana. Plant Cell 22: 4084–4097

Tholl D, Lee S (2011) Terpene Specialized Metabolism in Arabidopsis thaliana. The Arabidopsis Book 9: e0143

Tran L-SP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2004) Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. Plant Cell 16: 2481–2498

Trontin C, Tisné S, Bach L, Loudet O (2011) What does Arabidopsis natural variation teach us(and does not teach us) about adaptation in plants? Curr Opin Plant Biol 1–7

Tsai HL, Li YH, Hsieh WP, Lin MC, Ahn JH, Wu SH (2014) HUA ENHANCER1 is involved in posttranscriptional regulation of positive and negative regulators in Arabidopsis photomorphogenesis. Plant Cell 26: 2858–2872

Tsukaya H (2002) Interpretation of mutants in leaf morphology: genetic evidence for a compensatory system in leaf morphogenesis that provides a new link between cell and organismal theories. International review of cytology. doi: 10.1016/S0074-7696(02)17011-2

Ueno Y, Ishikawa T, Watanabe K, Terakura S, Iwakawa H, Okada K, Machida C, Machida Y (2007) Histone deacetylases and ASYMMETRIC LEAVES2 are involved in the establishment of polarity in leaves of Arabidopsis. Plant Cell 19: 445–457

Valdés AE, Övernäs E, Johansson H, Rada-Iglesias A, Engström P (2012) The homeodomain-leucine zipper (HD-Zip) class I transcription factors ATHB7 and ATHB12 modulate abscisic acid signalling by regulating protein phosphatase 2C and abscisic acid receptor gene activities. Plant Mol Biol 80: 405–418

Van Houtte H, Vandesteene L, López-Galvis L, Lemmens L, Kissel E, Carpentier S, Feil R, Avonce N, Beeckman T, Lunn JE, et al (2013) Overexpression of the trehalase gene AtTRE1 leads to increased drought stress tolerance in Arabidopsis and is involved in abscisic acid-induced stomatal closure. Plant Physiol 161: 1158–1171

Van Landeghem S, Hakala K, Rönnqvist S, Salakoski T, Van de Peer Y, Ginter F (2012) Exploring biomolecular literature with EVEX: connecting genes through events, homology, and indirect associations. Adv Bioinformatics 2012: 582765

Van Landeghem S, De Bodt S, Drebert ZJ, Inzé D, Van de Peer Y (2013) The potential of text mining in data integration and network biology for plant research: a case study on Arabidopsis. Plant Cell 25: 794–807

van Leeuwen H, Kliebenstein DJ, West MAL, Kim K, van Poecke R, Katagiri F, Michelmore RW, Doerge RW, St Clair DA (2007) Natural variation among Arabidopsis thaliana accessions for transcriptome response to exogenous salicylic acid. Plant Cell 19: 2099–2110

Vandesteene L, Lopez-Galvis L, Vanneste K, Feil R, Maere S, Lammens W, Rolland F, Lunn JE, Avonce N, Beeckman T, et al (2012) Expansive evolution of the TREHALOSE-6-PHOSPHATE PHOSPHATASE gene family in Arabidopsis. Plant Physiol 160: 884–896

Vanhaeren H, Gonzalez N, Coppens F, De Milde L, Van Daele T, Vermeersch M, Eloy NB, Storme V, Inzé D (2014) Combining growth-promoting genes leads to positive epistasis in Arabidopsis thaliana. eLife 3: e02252

Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 7: 91

Verslues PE, Sharma S (2010) Proline Metabolism and Its Implications for Plant-Environment Interaction. The Arabidopsis Book 8: e0140

Vieira Dos Santos C, Rey P (2006) Plant thioredoxins are key actors in the oxidative stress response. Trends Plant Sci 11: 329–334

Vile D, Pervent M, Belluau M, Vasseur F, Bresson J, Muller B, Granier C, Simonneau T (2011) Arabidopsis growth under prolonged high temperature and water deficit: independent or interactive effects? Plant Cell Environ 35: 702–718

Vilhjálmsson BJ, Nordborg M (2012) The nature of confounding in genome-wide association studies. Nature Publishing Group 14: 1–2

Wang S, Blumwald E (2014) Stress-induced chloroplast degradation in Arabidopsis is regulated via a process independent of autophagy and senescence-associated vacuoles. Plant Cell 26: 4875–4888

Wayne ML, Pan YJ, Nuzhdin SV, McIntyre LM (2004) Additivity and trans-acting effects on gene expression in male Drosophila simulans. Genetics 168: 1413–1420

West MAL, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, St Clair DA (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. Genetics 175: 1441–1450

White DWR (2006) PEAPOD regulates lamina size and curvature in Arabidopsis. Proc Natl Acad Sci USA 103: 13238–13243

Wright IJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, Cavender-Bares J, Chapin T, Cornelissen JHC, Diemer M, et al (2004) The worldwide leaf economics spectrum. Nature 428: 821–827

Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26: 873–881

Wu Y, Cosgrove DJ (2000) Adaptation of roots to low water potentials by changes in cell wall extensibility and cell wall proteins. J Exp Bot 51: 1543–1553

Xu Z, Zhou G (2008) Responses of leaf stomatal density to water status and its relationship with photosynthesis in a grass. J Exp Bot 59: 3317–3325

Yang H, Matsubayashi Y, Nakamura K (2001) Diversity of Arabidopsis genes encoding precursors for phytosulfokine, a peptide growth factor. Plant Physiol

Yoo CY, Pence HE, Jin JB, Miura K, Gosney MJ, Hasegawa PM, Mickelbart MV (2010) The Arabidopsis GTL1 transcription factor regulates water use efficiency and drought tolerance by modulating stomatal density via transrepression of SDD1. Plant Cell 22: 4128–4141

Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L (2003) Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat Genet 35: 57–64

Zemach A, Grafi G (2003) Characterization of Arabidopsis thaliana methyl-CpG-binding domain (MBD) proteins. Plant J 34: 565–572

Zerial M, McBride H (2001) Rab proteins as membrane organizers. Nat Rev Mol Cell Biol 2: 107–117

Zhang C-J, Ning Y-Q, Zhang S-W, Chen Q, Shao C-R, Guo Y-W, Zhou J-X, Li L, Chen S, He X-J (2012) IDN2 and its paralogs form a complex required for RNA–directed DNA methylation. PLoS Genet 8: e1002693–14

Zhang J, Nguyen HT, Blum A (1999) Genetic analysis of osmotic adjustment in crop plants. J Exp Bot 50: 291–302

Zhao J, Favero DS, Peng H, Neff MM (2013) Arabidopsis thaliana AHL family modulates hypocotyl growth redundantly by interacting with each other via the PPC/DUF296 domain. Proc Natl Acad Sci USA 110: E4688–E4697

Zourelidou M, de Torres-Zabala M, Smith C, Bevan MW (2002) Storekeeper defines a new class of plant-specific DNA-binding proteins and is a putative regulator of patatin expression. Plant J 30: 489–497

Conclusions and future directions

Drought stress has a tremendous effect on plants. Many plant traits such as organ growth, flowering time, general morphology and metabolism are affected by drought. The selection pressure that is imposed by drought led to many structural and molecular adaptations. Most papers dealing with drought stress, study the adaptations of plants to severe stress conditions, often threatening survival of the plant. However, when plants are exposed to relatively mild drought conditions they show a precautionary growth arrest of which the basis is poorly studied.

In this thesis we set out to investigate whether the mechanisms that are involved in protecting plants to severe drought stress are also involved in mild drought stress. We found that none of the screened mutants that previously showed improved tolerance to severe stress also performed better in mild drought stress. We concluded that different mechanisms are active upon mild and severe drought stress and that the response to stress depends on the severity or dosage. Because we are particularly interested in the growth reduction upon drought, this thesis focuses on mild drought responses, which typically reduce growth without threatening the survival of the plant. More specifically we aimed to unravel the effect on leaf growth, a highly complex trait that involves a great number of genes (Gonzalez et al., 2012; Kalve et al., 2014). By adapting to different environmental conditions, such as differences in water availability, natural Arabidopsis accessions show substantial genetic variation in the genes and regulatory networks that underlie specific phenotypes. Hence, natural variation offers an excellent tool to screen for the genetic mechanisms that are involved in complex traits.

A pilot study on six accessions revealed distinct responses to mild drought of the different leaf growth-related phenotypes. The transcriptome analysis of young developing leaves identified 354 genes that showed a response to mild drought stress that was conserved in the different accessions. This set of genes contained different components of the core ABA signaling machinery such as ABA receptors (PYL/PYR/RCARs), protein phosphatases (PP2Cs) and ABA responsive element binding factors (ABFs). But also genes downstream of the ABA signaling cascade such as a proline synthase (P5CS1) and oxidase (ERD5). The ABA-related genes were largely overlapping with responses to the severe drought response and the mild drought response of mature tissue and are therefore part of a general drought response instead of being specific for the mild drought response of young developing leaves. Another large functional group of genes that showed a common transcription response are involved in cell wall modification. The up-regulation of the cell wall loosening expansins and pectin lyases in young developing leaf tissue differed from the response of mature tissue to severe and mild drought. The loosening of the cell wall allows maintaining the growth at reduced turgor and/or promoting an earlier onset of cell expansion. In conclusion, drought, independent of its severity, activates a number of mechanisms such as ABA signaling. Nonetheless, mechanisms like cell wall loosening show a specific response to mild drought stress in growing leaf tissue. The drought response thus likely contains a core response with additional context-dependent fine-tuning.

To further detect genes that play a role in determining the responses of the different leaf growth-related phenotypes to mild drought stress a GWAS analysis was conducted. Therefore, different leaf growth related phenotypes were measured in a set of 98 accessions. Phenotypes ranged from

rosette and leaf area measurements to more detailed cellular characteristics such as cell number, cell area and stomatal index. All phenotypes showed substantial natural variation and were significantly affected by the mild drought treatment. Accessions that were large in control conditions tended to be large in mild drought as well, but were not more sensitive to the stress than small accessions.

Because of the relatively low amount of accessions screened and the high density of the genotypic markers, we did not have sufficient statistical power to detect statistical significant associations. Nevertheless, interesting candidate genes for follow-up studies were retrieved: *mi*R*171c*, *CSLC4*, *SAP12* and *EXL1*. Three of the candidate genes (*CSLC4*, *SAP12*, *EXL1*) are also differentially expressed upon mild drought, averaged over 89 accessions.

In order to identify the set of genes that can predict whether a sample was subjected to mild drought or control conditions a modeling approach was conducted, using the genome-wide transcriptome data of 89 accessions. An ANOVA analysis delivered 283 differentially expressed genes over the different accessions. This set of 283 genes was successfully used in a classification model to predict whether a sample was subjected to control or mild drought conditions. As expected, this set of 283 genes is enriched for genes involved in drought responses and growth regulation. The stress predictors that are not yet described as involved in growth regulation upon drought stress are interesting candidates for further investigation of a putative role in the mild drought response of young developing leaves.

The substantial variation in gene expression between the different accessions was used in the eGWAS. This delivered putative regulatory interactions between gene expression and specific genetic loci. From the eGWAS it seemed that trans regulation is playing an important role in the response to mild drought stress. This confirms the general idea (Kliebenstein, 2009; Cubillos et al., 2012) that trans regulatory elements are important for the expression upon specific conditions (environment, cell type, developmental stage), whereas *cis* regulatory elements are thought to be more important in regulating gene expression in general. The *trans* loci that were detected in regulating the differential expression upon mild drought stress covered a wide range of functions such as transcription factors, hormone biosynthesis, epigenetic regulation and transposable elements.

The different genes and regulators that have been proposed in this thesis to be important for the mild drought response of plants are interesting candidates to further screen for orthologous genes in crops. Such orthologs are interesting candidates for engineering for more drought tolerant plants. However, I think the main interest of this thesis towards applications is that we have shown the possibility to use natural variation in plants to gain more insight in the genetic architecture of specific traits under different environmental conditions. The approaches suggested here, in combination with the future possibilities that are discussed below, will be valuable tools to use in crops in order to detect their specific genetic architecture.

Future directions

In this thesis we used natural variation to unravel mild drought responses of growth-related phenotypes. Although we already obtained some clues on genes and processes that are involved in the response, to our opinion, the presented results are just the tip of the iceberg. The dataset that was produced is dense in information and further analyses are required to uncover the full potential of this data. Therefore a few suggestions are postulated in this final part of the thesis.

The phenotypes that were measured in this thesis give an overall idea on leaf growth responses upon mild drought. However, it would be interesting to do additional physiological measurements to better understand the behavior of the different accessions. For example, the transcriptome analyses suggest a main role for ABA in this response. ABA is known to have antagonistic regulatory interactions with GA and ethylene (Gómez-Cadenas et al., 2001; Cheng et al., 2009). Therefore it will be interesting to measure the concentrations of these different hormones in order to analyze how their levels and their mutual ratios vary between the different accessions in response to mild drought. The different accessions also showed an upregulation of the proline biosynthesizing P5CS1. Measuring the actual proline levels can therefore highlight the variation between the accessions in proline accumulation. The different hormone and proline measurements will allow for correlations with the leaf growth and cellular parameters and shed light on their regulation.

In this thesis, tolerance and sensitivity upon drought was determined as better or worse growing behavior under mild stress. Since this observation does not say anything about growth efficiency in drought, it would be useful to measure, in the different accessions, the water use efficiency (WUE) which is a measure for the amount of carbon that is assimilated per quantity of water.

Concerning the performed GWAS in this thesis, we realized that we clearly lack power to significantly associate phenotype and genotype. In order to increase the statistical power of the GWAS that links genotype to phenotype, the number of accessions can be increased. It has been shown that screening twice the amount of accessions can result in a more than doubling of the statistical power (Supplementary Figure 1; Korte and Farlow, 2013). The GWAS was performed with a very dense set of SNPs (±4 million), causing a stringent multiple testing correction. When the goal of a GWAS is to find the region where a causal SNP is located and not to define the causal SNP itself, not all SNPs need to be tested. Due to linkage, SNPs that are closely related to the causal SNP often associate with the phenotype as well. In order to define the region where the causal SNP can be found, a less dense subset of SNPs can be used. Crucial in this approach is the selection of the SNP subset. This subset can be a random subset of SNPs, but it is more strategic to select representative SNPs for each region with high linkage disequilibrium (LD). In a sliding window approach, the correlation between SNPs can be calculated per window. SNPs that are in a region of high LD are correlated and can be removed, except for one representative SNP, in order to obtain a less dense but valuable subset of SNPs. This methodology to reduce the number SNPs is called SNP pruning and can be performed with the PLINK tool (Purcell et al., 2007). The associations that are detected in the initial GWAS that makes use of the SNP subset will indicate the regions where the causal SNP is located. In order to retrieve the causal SNP, a targeted GWAS is performed specifically on the regions that are depicted from the initial GWAS, with all available SNP data for those regions. It will need to be checked to what extent the ±4

million tested SNPs can be reduced in number by this approach, but if even only one in two SNPs are removed due to correlation by SNP pruning, this already halves the number of tests that need to be performed in the initial GWAS. Suppose we find 5 regions that require further investigation of the 10kb region. Under the assumption that there are 300 SNPs in an average 10kb region (corresponds to 4 million SNPs on a total genome size of 135 Mbp), this will add 1500 extra tests to the stepwise GWAS, which brings the total to 2,001,500 tests. In comparison, a GWAS on the full SNP dataset would perform 4 million tests. A stepwise approach thus greatly reduces the amount of tests, resulting in a less stringent multiple testing correction and leaving sufficient statistical power to perform a detailed GWAS on the relevant regions.

Whereas the GWAS links genotype and phenotype, the eGWAS further digs into the transcriptional regulation of a gene. In this thesis the selection of the genes for the eGWAS was based on heritability. Because of the strong genetic component that explains expression variation of the selected genes, this criterion increased the chance to detect a locus that associates with gene expression. However, it does not necessarily mean that these selected genes are involved in the mild drought response of young developing leaves. To get a broader view on the transcriptional regulation to mild drought stress, it will be useful to perform the eGWAS on genes that are selected for their transcriptional responses to the mild drought response. The stress predictors are in this respect excellent candidates, but also genes identified by GWAS can be screened for their transcriptional regulators by eGWAS. A selection of more biologically relevant genes for the eGWAS will definitely deliver more insight in the transcriptional regulation of the mild drought response of young developing leaves.

Model inflation led to the exclusion of a large number of genes from the eGWAS analysis. It is likely that also a number of genes for which the model was not inflated, were excluded due to the stringent selection. In retrospect, the selection could be made less stringent, and data from other analyses/sources (GWAS, co-expression, literature, etc.) should then be added as extra evidence for regulatory interactions. Moreover, the integration of different data sources will eventually allow for the construction of a gene regulatory network (GRN). GWAS data on a phenotype that is closely related to the tissue for which the transcriptome is characterized will deliver good candidate genes for the eGWAS. In addition, the regulators that are retrieved in the eGWAS and that also associate to the phenotype are potential important nodes in the GRN. Co-expression data can be added to the network as suggested co-regulated clusters. Also text-mining tools for large-scale literature screens, such as EVEX (Van Landeghem et al., 2013), can deliver extra evidence for the regulatory interactions in the GRN. In this integrated approach some nodes and edges from the network will be suggested by one data source, while others will have evidence form multiple sources. A weight of evidence will need to be calculated for each of the interactions in the network, based on the amount and the source of evidence. For example, a regulatory interaction suggested by the eGWAS with evidence from literature and with both genes associating to the phenotype should receive a higher weight than an interaction suggested only by the eGWAS. Because published results may proof the causality of a regulatory interaction, they offer crucial information for constructing the GRN. However, we realize that the search for new candidate genes might be hampered by the fact that literature evidence will create a bias towards the known interactions, pushing unknown but potentially interesting interactions to the background. Therefore it will be practical to have besides the weight of evidence, which indicates how plausible a certain interaction is, also a factor that indicates

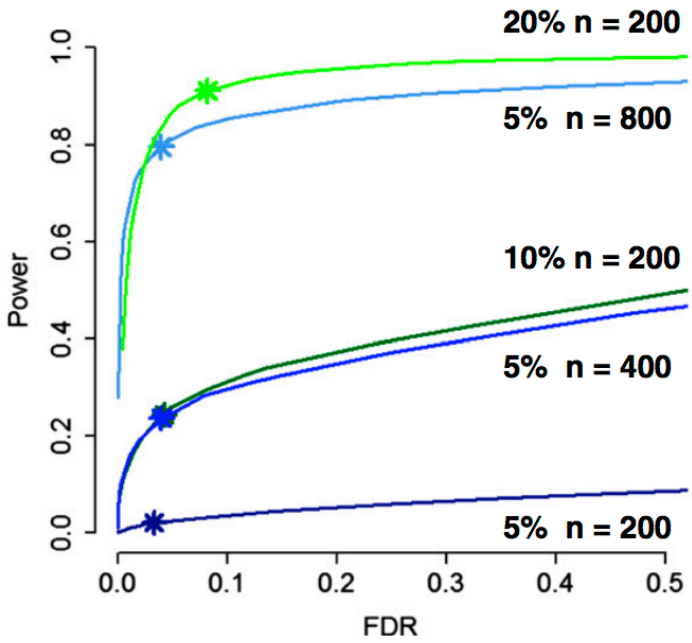
whether an interaction is proven or not. This allows for giving a lower weight to evidence from literature, in order to avoid a bias towards known interactions. While the information on proven interactions still adds full evidence in constructing the network when the factor of proven interactions is taken into account. If the scope of the research is to find a robust and high confident GRN, the interactions can be selected on this factor of proven interactions and a very stringent weight cut-off. While the search for new candidates can rely on a more liberal selection, based on weight only.

Once the GRN is established it is possible to define the effects of the different polymorphisms in the different accessions on the performance of the network. Structural equation models (SEMs), where the expression of each gene is determined by the expression of upstream regulators and their regulatory effect on the downstream gene, are suggested to further unravel the performance of GRNs in natural variation (Nuzhdin et al., 2012). In different natural variants, the regulatory effect between nodes is allele-specific, resulting in allele-specific expression of the downstream nodes. Because we have the allele information and the transcriptome data of different accessions, the allele-specific expression can be calculated. A SEM can then identify the polymorphisms that change the performance of the GRN and thus the genetic changes that are required to specifically engineer the GRN. Furthermore, by measuring the allele-specific expression of the homologs in another species, e.g. a crop, the SEM can validate or further fine-tune the GRN in a non-model organism. Similarly, a Bayesian approach has been suggested to detect the specific effect of certain polymorphisms on the performance of the GRN and to infer which GRN perturbations are causing certain phenotypic effects (Marjoram et al., 2013).

In this thesis we focused, with the GWAS, on the link between genotype and phenotype and, with eGWAS, on the link between genotype and gene expression. A similar GWAS-like approach can, in principle, be used to link the extensive variation in gene expression to the phenotypes. The association between transcriptome and phenotype will directly deliver the genes that are transcriptionally involved in determining the phenotype. Some genes that are not detected in GWAS, because they lack genetic polymorphisms that correlate with the phenotype due to conservation, but that do differ in expression levels between accessions will be retrieved by mapping the phenotype to the transcriptome. Knowing the genes that are transcriptionally involved in a specific phenotype is important for constructing the underlying GRN.

Finally, we have shown that natural variation is a very powerful tool to learn more about the adaptations of different accessions to specific conditions. The large genetic variation among the different Arabidopsis populations can be exploited to uncover how plants cope with adverse conditions, such as drought stress. With GWAS we have a tool at hand to study the genes and regulatory mechanisms that determine the phenotype. I am convinced that by integrating different GWAS-like approaches on genotype, transcriptome, phenotype, but also metabolome, proteome and interactome data will deliver insight in the underlying regulatory network of specific traits and their adaptations to different environmental conditions. This information will allow modelers to construct GRNs and predict the phenotypic effect of specific perturbations in the network. When this approach will be conducted on crop species it will, to my opinion, identify the alleles at the nucleotide level that are responsible for a specific phenotypic response upon a particular environmental perturbation. Once these alleles are known, crops can be engineered to cope with the challenges that agriculture is facing in terms of climate change and limited resource availability.

Supplementary figures



Supplementary Figure 1: Power and FDR for an idealized phenotype. Simulations in which a single random SNP explaining 5%, 10% or 20% of the phenotypic variance (with heritability ~0.75) were performed in either 200, 400 or 800 individuals. Simulations are based on the available SNP data for Arabidopsis [20], with structure added by giving 10,000 random SNPs a tiny effects size. The star indicates power (the ability to find true positives) and FDR (false positives) at the 5% bonferroni-corrected threshold for 220,000 markers (Figure derived from Korte and Farlow, 2013).

References

Cheng W-H, Chiang M-H, Hwang S-G, Lin P-C (2009) Antagonism between abscisic acid and ethylene in Arabidopsis acts in parallel with the reciprocal regulation of their metabolism and signaling pathways. Plant Mol Biol 71: 61–80

Cubillos FA, Coustham V, Loudet O (2012) Lessons from eQTL mapping studies: non-coding regions and their role behind natural phenotypic variation in plants. Curr Opin Plant Biol 15: 192–198

Gómez-Cadenas A, Zentella R, Walker-Simmons MK, Ho T-HD (2001) Gibberellin/Abscisic Acid Antagonism in Barley Aleurone Cells: Site of Action of the Protein Kinase PKABA1 in Relation to Gibberellin Signaling Molecules. Plant Cell 13: 667–679

Gonzalez N, Vanhaeren H, Inzé D (2012) Leaf size control: complex coordination of cell division and expansion. Trends Plant Sci 17: 332–340

Kalve S, De Vos D, Beemster GTS (2014) Leaf development: a cellular perspective. Front Plant Sci 5: 362

Kliebenstein D (2009) Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. Annu Rev Plant Biol 60: 93–114

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9: 29

Marjoram P, Zubair A, Nuzhdin SV (2013) Post-GWAS: where next? More samples, more SNPs or more biology? Heredity 112: 79–88

Nuzhdin SV, Friesen ML, McIntyre LM (2012) Genotype-phenotype mapping in a post-GWAS world. Trends in Genetics 28: 421–426

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al (2007) PLINK: A tool Set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81: 559–575

Van Landeghem S, De Bodt S, Drebert ZJ, Inzé D, Van de Peer Y (2013) The potential of text mining in data integration and network biology for plant research: a case study on Arabidopsis. Plant Cell 25: 794–807

Acknowledgements

This thesis and the entire PhD would never have been possible without the help and support of many other people. Therefore I want to thank all these people here.

I want to thank Dirk, for giving me the opportunity to do this research and taking the risk to step into the complex world of natural variation. It always amazed me that despite your extremely busy agenda, you always found/made time to read my writings and discuss the results.

Frederik and Nathalie, you guys actually deserve a statue. You learned me a lot about many aspects of science. Knowledge and skills I will definitely use throughout my entire life.

I very much appreciate it that both of you were constantly challenging me and taught me in this way to be very critical about my own work. Nathalie, I want to specially thank you to threaten me every once in a while to kick my ass. Frederik, I am very grateful that you were guiding me into bioinformatics. To many of the problems I faced you could reach me a solution and for many of the errors in my scripts you were able to find the cause.

Jasper, Hannes, Stefan, (Michiel) although you all left my alone the last year, the years before it was always relaxing to take a break with you and discuss about practically everything.

Twiggy, Liesbeth, Katrien and Mattias, you are all doing a great job by conducting a lot of work in the group. Without your help big experiments like screening a bunch of accessions would never be possible.

The entire Yield group for all the fun moments and nice atmosphere.

Kristof Verleye for cutting the WIWAM pots, always with the smile on your face.

The people from bioinformatics, I want to thank you for accepting this strange wetlab creature that still smelled like potsoil, in your room. The help with, often basic, informatics problems is really appreciated, as were the absurd coffee break conversations.

Vader en moeder, merci om me al de kansen te geven die dit mogelijk gemaakt hebben. Jullie zijn de basis van dit alles.

Lotte, te midden alle hectiek van het schrijven leerden we elkaar kennen. Een belangrijk deel van deze thesis is dan ook geschreven in een trein ergens tussen Gent en Antwerpen. Merci om na die treinritten telkens voor de nodige ontspanning te zorgen.

Bram en Marijke, merci voor het geduld met de bergen afwas die een roeiende doctorandus produceert. Het was zeer fijn om met jullie samen te wonen. En weet: "Waarheen we ook gaan, team fuchsia zal altijd blijven bestaan!".

De biologen, voor alle trappisten, weekends, kotentochten …

De GRS, omdat een gezonde geest nog altijd best in een gezond lichaam vertoeft. Daar bestaat zelfs een wetenschappelijk formule voor. Het deed me dikwijls enorm veel deugd om na een dag in het labo op de watersportbaan te mogen afzien.

Last but not least I also want to thank the IWT for their financial support and CropDesign (BASF) for the collaboration that allowed for producing the majority of the results in this thesis.

To all whom I might have forgotten, Thank You!